A Data-Driven Toolchain for the Operational Performance Analysis and Optimization of Buildings

by

Anthony Reed Florita

B.S., University of Wyoming, 2004

M.S., University of Nebraska–Lincoln, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Civil, Environmental and Architectural Engineering

2019

This thesis entitled:
A Data-Driven Toolchain for the Operational Performance Analysis and Optimization of Buildings
written by Anthony Reed Florita
has been approved for the Department of Civil, Environmental and Architectural Engineering

Prof. Gregor P. Henze, Ph.D., P.E.

Prof. Michael J. Brandemuehl, Ph.D., P.E.

Prof. Balaji Rajagopalan, Ph.D.

Asst. Prof. Kyri A. Baker, Ph.D.

Robert H. Dodier, Ph.D.

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Florita, Anthony Reed (Ph.D., Architectural Engineering)

A Data-Driven Toolchain for the Operational Performance Analysis and Optimization of Buildings

Thesis directed by Prof. Gregor P. Henze, Ph.D., P.E.

Automation has proven indispensable to advancing human endeavors. Within the built environment its evolution and sophistication are on the cusp of moving beyond automatic control into automated prediction and diagnosis. A data-driven toolchain is developed so human efforts can be focused on high-value concerns. The research examines smart buildings as a cyberphysical construct and places the Bayesian perspective as paramount. Prior knowledge is leveraged through common building energy modeling and simulation tools, which are utilized and extended. An iterative, three-step process is developed to 1) classify building energy performance scenarios, 2) forecast dynamics over a planning horizon of interest, and 3) signal human decision-makers concerning deviations from ideal behavior. In the classification step, focus is placed on the discrete wavelet transformation of electrical demand profiles, producing energy and entropy feature extraction from the wavelet levels at definitive time frames, and Bayesian probabilistic hierarchical clustering. The process yields a categorized and manageable set of representative electrical demand profiles for smart grid applications. In the forecasting step, a cyclical two-stage model predictive control process of policy planning followed by execution is evaluated. The results show that even the most complicated nonlinear autoregressive neural network with exogenous input does not appear to warrant the additional efforts in forecasting model development and training in comparison to the simpler models. In the signaling step, a simulation study is considered to assess whole-building energy signaling accuracy in the presence of uncertainty and faults at the submetered level, which may lead to tradeoffs at the whole-building level that are not detectable without submetering. Together, the steps form a data-driven toolchain for the operational performance analysis and optimization of buildings.

Dedication

To the dreamers, wanderers, and individualists...

Acknowledgements

I would like to express my deepest gratitude to my advisor Professor Henze for his continuous support during my research studies and in life, for his wisdom, drive, and patience. His guidance helped me surmount many obstacles. I cannot imagine having a better mentor.

In addition to my advisor, I would like to thank the rest of my committee: Professor Brandemuehl, Professor Rajagopalan, Dr. Baker, and Dr. Dodier, for their insightful feedback and encouragement. I am very grateful for their expert perspectives concerning my research, which helped expand my viewpoint and refocus my efforts when necessary.

My gratitude also goes to the Fraunhofer Institute for Solar Energy Systems, specifically Christian Neumann and Dirk Jacob, who provided me with a visiting scientist position and initiated much of my research. Without their invaluable support and deep understanding of building physics, my research would not have been possible.

Last but definitely not least, I would like to thank my family: you are my everything!

Contents

Chapter

1 Introduction 1

1.1 Background 1

1.2 Research Motivation 5

1.3 Organization of the Dissertation 9

2 Operational Building Energy Performance 11

2.1 Buildings in the Digital Age 11

2.2 Trends in Building Energy Modeling and Simulation 13

2.2.1 Sensitivity Analysis 15

2.3 Data-Driven Assessments 15

2.3.1 Typical Building Energy Performance Data 16

2.3.2 Calibration, System Identification, and Evidence-Based Approaches 17

2.4 Fault Detection and Diagnostics 25

2.4.1 Foundational Applications in Buildings 25

2.4.2 Building Primary and Secondary Systems 25

2.4.3 HVAC Control Systems 32

3 Development of a Data-Driven Toolchain 34

3.1 Research Statement 34

3.2 APEX Conceptualization 36

3.2.1 OpBEMs for Smart Buildings . . . 36
3.2.2 Tool Abstraction for One-of-a-Kind Buildings . . . 39
3.2.3 Building Health and Energy Performance . . . 40
3.3 The Piecewise Development of Tools . . . 41

4 Classification of Commercial Building Electrical Demand Profiles for Energy Storage Applications 44
4.1 Introduction . . . 44
4.2 Literature Review . . . 45
4.3 Methodology . . . 47
4.3.1 Discrete Wavelet Transformation . . . 47
4.3.2 Feature Extraction . . . 51
4.3.3 Bayesian Probabilistic Hierarchical Clustering . . . 55
4.4 Results . . . 57
4.4.1 The Influence of Climate . . . 59
4.4.2 The Influence of Building Type . . . 60
4.4.3 The Influence of Building Size . . . 61
4.4.4 Perceived Use of the Commercial Building Classes . . . 63
4.4.5 A Classified Building Energy Storage Example . . . 64
4.5 Conclusions . . . 67

5 Comparison of Short-Term Weather Forecasting Models for Model Predictive Control 68
5.1 Introduction . . . 68
5.1.1 Nomenclature . . . 70
5.2 Literature Review . . . 71
5.3 Description of Analysis . . . 74
5.3.1 Overview of Short-Term Weather Prediction . . . 74

5.3.2 Investigated Forecasting Algorithms 77
5.3.3 Neural Networks (NN) 84
5.4 Results and Discussion 87
5.4.1 Performance Comparison 89
5.4.2 Discussion 91
5.5 Conclusions 95

6 An Energy Signal Tool
for Decision Support in Building Energy Systems 97
6.1 Introduction and Motivation 97
6.1.1 Nomenclature 101
6.2 Literature Review 105
6.3 Methodology 107
6.3.1 Modeling Environment 107
6.3.2 Uncertainty Quantification 125
6.3.3 Decision Analysis 130
6.3.4 Bayesian Updating 135
6.4 Results 141
6.4.1 Decision Support Case Studies 141
6.4.2 Bayesian Parameter Updating Case Studies 149
6.5 Summary and Conclusions 161

7 Conclusions and Future Research 162
7.1 Overview 162
7.2 Uncertainty Quantification 165
7.3 Future Research 167

Bibliography 168

Appendix

A A Summary of the Model-Based Benchmarking (ModBen) Project 183

B Fundamentals of Bayesian Inference 187

B.1 Derivation of Bayes' Theorem . . . 187

B.2 Bayesian Parameter Uncertainty Quantification . . . 189

B.3 Bayesian Model Comparison . . . 191

B.4 Requirements for Computation . . . 192

B.4.1 Integral Estimation via Monte Carlo Method . . . 192

B.4.2 Markov Chain Monte Carlo (MCMC) . . . 194

B.4.3 Metropolis Algorithm . . . 194

B.4.4 Metropolis-Hastings Algorithm . . . 195

B.4.5 Gibbs Sampler . . . 195

C Identifying Wind and Solar Ramping Events 197

C.1 Introduction . . . 197

C.2 Swinging Door Algorithm . . . 199

C.3 Wind and Solar Data . . . 204

C.4 Results . . . 206

C.5 Conclusions . . . 211

Tables

Table

2.1 Common faults found in chillers. . . . 26

2.2 Heating system and boiler faults common to practice. . . . 27

2.3 Common faults in AHU's air mixing section. . . . 28

2.4 AHU's filter-coil section common faults: Part 1 of 3. . . . 29

2.5 AHU's filter-coil section common faults: Part 2 of 3. . . . 30

2.6 AHU's filter-coil section common faults: Part 3 of 3. . . . 30

2.7 Faults common to an AHU's fan section. . . . 31

2.8 Common VAV faults: Part 1 of 2. . . . 33

2.9 Common VAV faults: Part 2 of 2. . . . 33

5.1 Geographic Locations Considered for Short-Term Weather Prediction. . . . 76

5.2 Labeling of the Various Models Used in the Analysis. . . . 78

5.3 Forecasting Models' Mean %CV over all Locations for each Weather Variable Predicted. . . . 88

5.4 Forecasting Models' Mean %MBE over all Locations for each Weather Variable Predicted. . . . 88

6.1 Selected EnergyPlus Model Details. . . . 109

6.2 Model complexity results. . . . 116

6.3 Calibrated Five-Parameter Network RC Parameters. . . . 120

6.4 Retail Building Use Parameters. . . . 126

6.5 Retail Building HVAC Parameters. . . . 126

6.6 Fault Ranges. 127

6.7 Summary Statistics of Conditional Distributions of Whole-Building, HVAC, Lighting, and Plug Electricity Consumption in [%] in [MWh/a]. 130

6.8 Decision Analysis State and Action. 133

6.9 Building Model Truth Parameters. 151

Figures

Figure

1.1 Conceptualization of a smart grid: the next generation electric grid for enhanced communication, renewable energy integration, and whole-system efficiency. 4

1.2 Building Automation System (BAS) topology: typical hierarchy for building control. 5

1.3 The Plan-Do-Study-Act (PDSA) method (or Deming cycle) is an iterative four-step management method used to control and continuously improve processes and products. 8

2.1 Typical multiple-input, multiple-output (MIMO) system identification. Various algorithms are used to tune the model according to its prediction error. 20

2.2 The system identification procedure can be roughly broken into six stages, with the final, seventh validation stage used for accepting or revising the model as necessary; adapted from Nelles (2000) and modified according to a Bayesian perspective. 20

3.1 A Posteriori EXplorer (APEX) system concept for buildings. 37

3.2 The three-step process envisioned as a spiral graph: continual improvement is made in process control by the classification, forecasting, and signaling of operational performance analysis. 43

4.1 The Discrete Wavelet Transformation (DWT) performed using a filter bank for multiresolution analysis of the electrical demand profiles. 50

4.2 The first three levels of a DWT using the Daubechies-4 wavelet: the magnitude at a given time indicates the correlation strength at that level, units are not included for visual clarity. 51
4.3 Two typical building electrical demand profiles and their relative wavelet energy comparison. 53
4.4 Frequency plot of total wavelet entropy for each of the 4,820 commercial buildings. . 54
4.5 Dendrogram of the clustering process for the united states commercial building stock: visualization of the 4,820 buildings merging into the final 114 clusters (building classes) according to the 25 features extracted from the electrical demand profiles; for visual clarity building labels not included. 58
4.6 Limited portion of dendrogram (with focus on clusters 8 and 9) demonstrating merging hierarchy; red dashed lines are merges not made, while the number displayed indicated log odds for merging. 59
4.7 Clustering heatmap based on united states climate zones: CBECS uses five climate regions based on cooling and heating degree days. Darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number. 61
4.8 Clustering heatmap based on principal building activity: CBECS classifies buildings according to their major usage, e.g. office, retail, etc. darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number. 62
4.9 Clustering heatmap based on square footage. Darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number. 62
4.10 The building demand profile, on-site PV generation, and difference for the month of July. 65
4.11 Energy spillage and peak demand savings as a function of energy storage parameters. 66

5.1 Cyclical two-stage process of planning and execution horizon. 80
5.2 Recurrent neural network architecture. 86
5.3 Multiple comparisons using CV. 92
5.4 Multiple comparisons using MBE. 92
5.5 Forecast prediction in the face of simple, nearly sinusoidal weather variation. 94
5.6 Forecast prediction in the face of complex day-to-day variations in weather. 95
5.7 Forecast prediction in the face of highly uncertain, erratic weather variation. 95

6.1 Energy signal tool flowchart. 99
6.2 Isometric view of five-zone retail building model. 108
6.3 Zone plan of five-zone retail building model. 108
6.4 Twenty-one-parameter thermal RC network. 111
6.5 Eighteen-parameter thermal RC network. 113
6.6 Thirteen-parameter thermal RC network. 113
6.7 Eleven-parameter thermal RC network. 114
6.8 Eight-parameter thermal RC network. 114
6.9 Five-parameter thermal RC network. 114
6.10 Packaged RTU. 121
6.11 Validation of packaged RTU SA dry bulb temperature. 122
6.12 Validation of packaged RTU SA humidity ratio. 122
6.13 Validation of packaged RTU RA dry bulb temperature. 122
6.14 Validation of packaged RTU RA humidity ratio. 122
6.15 Validation of packaged RTU electricity demand. 123
6.16 Validation of packaged RTU gas demand. 123
6.17 Retail sensible zone load comparison for NSU scenario. 124
6.18 Retail zone mean air temperature comparison for NSU scenario. 124
6.19 Retail sensible zone load comparison for precooling scenario. 124

6.20 Retail zone mean air temperature comparison for precooling scenario. 124
6.21 Retail HVAC electricity consumption comparison for NSU scenario. 125
6.22 Retail HVAC electricity consumption comparison for precooling scenario. 125
6.23 Distributions of whole-building, HVAC, lighting, and plug electricity consumption. . 129
6.24 Relationship between deviation thresholds X_{low} and X_{high} and state probability vector $\vec{P}$. 132
6.25 Fault-free high consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 143
6.26 Fault-free medium consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 144
6.27 Fault-free low consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 145
6.28 Faulted high consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 146
6.29 Faulted medium consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 147
6.30 Faulted low consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure). 148
6.31 Example of four buildings managed by a building operator. 149
6.32 Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line. 151
6.33 Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line. 152

6.34 Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line. 152

6.35 DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line. 153

6.36 Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line. 153

6.37 Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line. 154

6.38 Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line. 154

6.39 Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line. 155

6.40 DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line. 155

6.41 Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line. 156

6.42 Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line. 156

6.43 Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line. 157

6.44 Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line. 157

6.45 DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line. 158

6.46 Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line. 158

C.1 Example of the piecewise linear approximation to a time series for ramp extraction and analysis; the scale is arbitrary for explanation purposes. 201

C.2 The swinging door algorithm for the extraction of ramps in power from the time series; the scale is arbitrary for explanation purposes. 201

C.3 Typical example of ramp extraction from two days of power at a large wind farm, showing up and down ramps of large, medium, and insignificant nature. 203

C.4 Typical example of ramp extraction from the first of two days of power at a PV solar plant, showing a clear day leading to a smooth profile. 204

C.5 Typical example of ramp extraction from the second of two days of power at a PV solar plant, showing up and down ramps because of clouds. 205

C.6 Bivariate distribution of wind power, ramp rise versus run, as a function of the ϵ value; top subplot is $\epsilon = 1\%$ maximum capacity in December, followed by $\epsilon = 2$, 3, and 5%, respectively. 207

C.7 Bivariate distribution of solar power, ramp rise versus run, as a function of the ϵ value; top subplot is $\epsilon = 1\%$ maximum capacity in December, followed by $\epsilon = 2$, 3, and 5%, respectively. 208

C.8 Bivariate distribution of wind power, ramp rise versus run, as a function of the aggregation level of wind turbines; top subplot is $\epsilon = 25\%$ total wind farm, followed by $\epsilon = 50$, 75, and 100%, respectively. An $\epsilon = 3\%$ of the maximum capacity was used for the month of December. 209

C.9 Bivariate distribution of solar power, ramp rise versus run, as a function of the aggregation level of PV modules; top subplot is $\epsilon = 25\%$ total plant, followed by $\epsilon = 50$, 75, and 100%, respectively. An $\epsilon = 3\%$ of the maximum capacity was used for the month of December 210

Chapter 1

Introduction

This chapter provides the background and motivation for why buildings are well-suited for automated prediction and diagnosis, both decision support mechanisms for operational building energy performance. There are, of course, major challenges to overcome. First, the built environment is a complex, ever-changing system. Second, any given building has peculiarities and special considerations necessary to classify and forecast its performance. Third, economic factors vary with time and space and must be integrated and harmonized with business processes and encompassing systems; i.e., the extent of automation should be optimized for signaling human decision-makers. After the establishment of a basis for research, the organization of the dissertation is given with research proposed toward: *a data-driven tool chain for the operational performance analysis and optimization of buildings.*

1.1 Background

According to present-day architectural vernacular: "buildings must become smart," with the phrase being open to countless interpretations. It is the modern-day adaptation to Le Corbusier's famous statement (c 1923) "A house is a machine for living in" [1]. Smart buildings could conjure the vision of a Jetson-esque futuristic utopia, complete with a robot maid who maintains technologies for every whimsical desire. Or, in contrast to robotic utility, they could enhance the occupants' health and sense of well-being through sustainable design and building health monitoring. What it takes to fully establish a smart building will change with time. Smart devices and services currently

exist and are being refined: from phones and televisions to cloud computing and the next generation electric grid. Fundamentally embodied in the term *smart* is the notion of increased functionality at a reduced effort. Computer hardware that once occupied entire building floors are now but tiny components of the omni-connected information age. In terms of software, data gathering that once required hours to ascertain (e.g., at a library) can now be queried and examined at the touch of a button. The envisioned communication amongst a network of smart devices and services has been termed the "Internet of Things" (IoT) [2] and, to some extent, is inevitable in the path toward smart buildings.

Ample opportunities exist for more intelligent building operations, yielding more satisfied occupants at enhanced energy performance. The built environment is responsible for approximately 41% of primary energy consumption in the United States, greater than transportation, 31%, and industry, 28%, with building site energy consumption dominated by heating (37% space; 12% water), cooling (10%), and lighting (9%) end-uses [3, 4]. Furthermore, it is estimated that 15–30% of commercial buildings' energy consumption is wasted due to poorly maintained, degraded, and improperly controlled equipment, necessitating fault detection, diagnostics, and prognostics (FDD&P) for building systems [5, 6]. Since World War II, energy consumption in the United States has grown from just over 30 quadrillion BTUs to just under 100 quadrillion BTUs — trending with industrialized nations and the requirements of an expanding population [3, 7].

The current United States primary energy production comes from approximately 77% fossil fuels, 11% nuclear, and 12% renewable generation [3]. The limited supply of hydrocarbons is debated [8] but the irreversible climate change due to carbon dioxide emissions is largely not [9]. Furthermore, the future of water supply is threatened by climate change [10]. This has profound consequences since, in 2010, thermoelectric power accounted for 45% of total water withdrawals, 38% of total freshwater withdrawals, and 51% of fresh surface-water withdrawals [11]. Reference case projections from 2010 to 2035 for the United States forecast energy consumption average annual growth rates of 0.3% (electrical demand growing by 0.7% per year, attributed primarily to the rising energy consumption in the buildings sector); average annual energy consumption growth

rates are 0.4% for natural gas and nuclear, 0.1% for coal, and 3.9% for non-hydropower renewable energy [12]. The variable and uncertain nature of renewable energy, and its expected share of total generation in 2035 ranging from 10-15% [12], indicate more will have to be done to accommodate its growth and, in general, electric power system planning, commitment, and dispatch [13]. More startling, world energy consumption is expected to increase by 53% between 2008 and 2035 [14], which will influence energy imports, exports, and prices.

Energy efficiency has been seen as a key means for achieving energy policy goals, yet there is a concern that continued improvements could be eliminated by increased energy consumption [15]. Furthermore, even for the most energy efficient buildings, power demands on the electric grid have high spatiotemporal correlation; i.e., regional demands are becoming more pronounced due to more than 80% of residents living in metropolitan areas of the United States [7]. Utilities have a formidable challenge in optimizing power flows on the electric grid, placing great strain on an aging infrastructure. From a holistic, system-of-systems perspective, it is possible the built environment is most energy efficient when buildings participate in managing the concerns of the electric grid; e.g., demand response, load shaping and modulation, or adaptive control according to renewable power availability. Building performance thus extends beyond energy efficiency into flexibility and supply following, with advanced analytics [13] used for integrating smart buildings and grids.

Major elements of the smart grid are illustrated in Figure 1.1; existing power infrastructure is shown with dashed lines to indicate decreasing centralization, while solid lines indicate an increasing decentralization. Expansion to the smart grid entails a more flexible and resilient grid which can more easily transfer power according to local supply-demand mismatch. Although the smart grid will be primarily centralized due to existing infrastructure, attributed to economies of scale in electric power generation, movement toward distributed generation is commonplace with rooftop photovoltaic (PV) systems; however, many types of renewable energy will sit alongside the existing infrastructure and participate according to availabilities and demands. Variability and uncertainty associated with renewable power generation, commonly termed "intermittency," requires buildings' flexibility while the game-theoretic agents of energy markets will determine dynamic pricing.

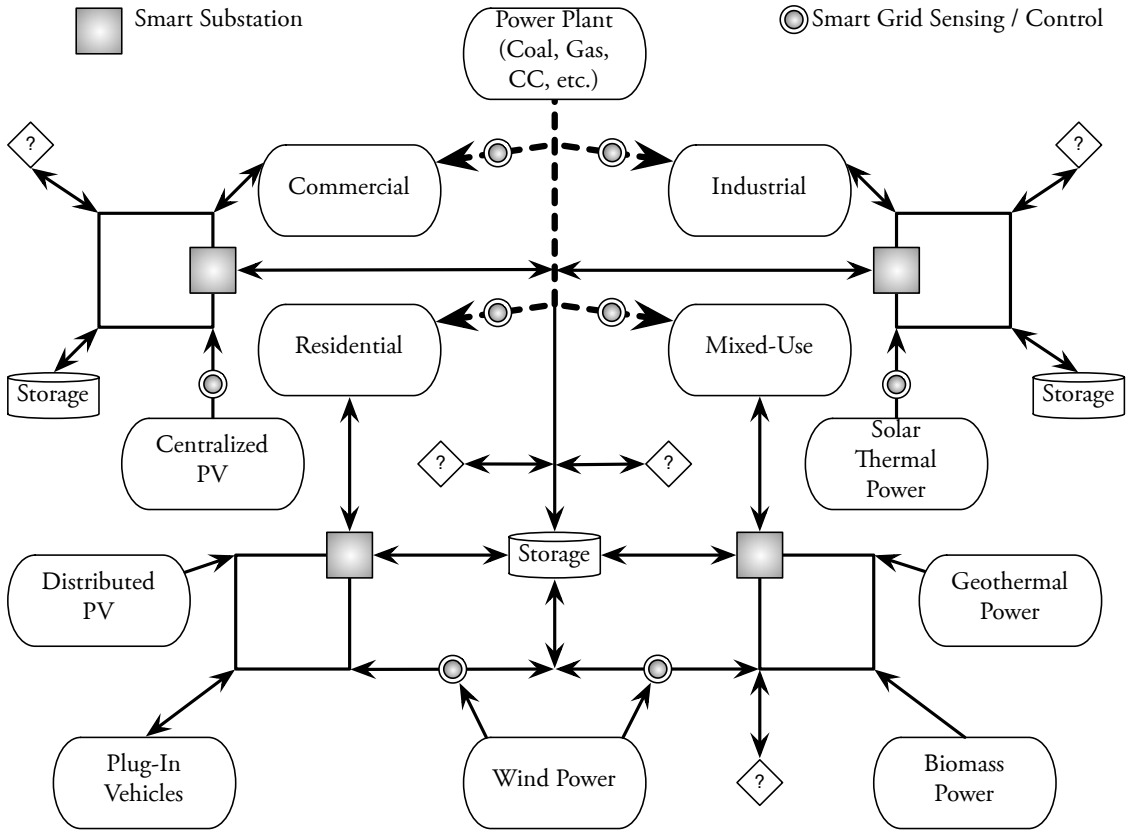

Figure 1.1: Conceptualization of a smart grid: the next generation electric grid for enhanced communication, renewable energy integration, and whole-system efficiency.

The hierarchy of a typical building automation system (BAS) is shown in Figure 1.2. It is primarily centralized, echoing contemporary electric power grid control. Minimal communication is enabled among elements within a given level of the system; i.e., horizontally in the figure. All necessary control can be accomplished in a top-down fashion; i.e., vertically in the figure. An energy management and control system (EMCS) acts as a further abstraction of the BAS for the observation and control of a building portfolio, and is accomplished through supervisory control and data acquisition (SCADA). Because this strategy mirrors that of the electric power grid, albeit at a much smaller scale, it is possible the systems will become indistinguishable with the emergence of smart buildings and grids — especially with publicly and privately owned IoT devices.

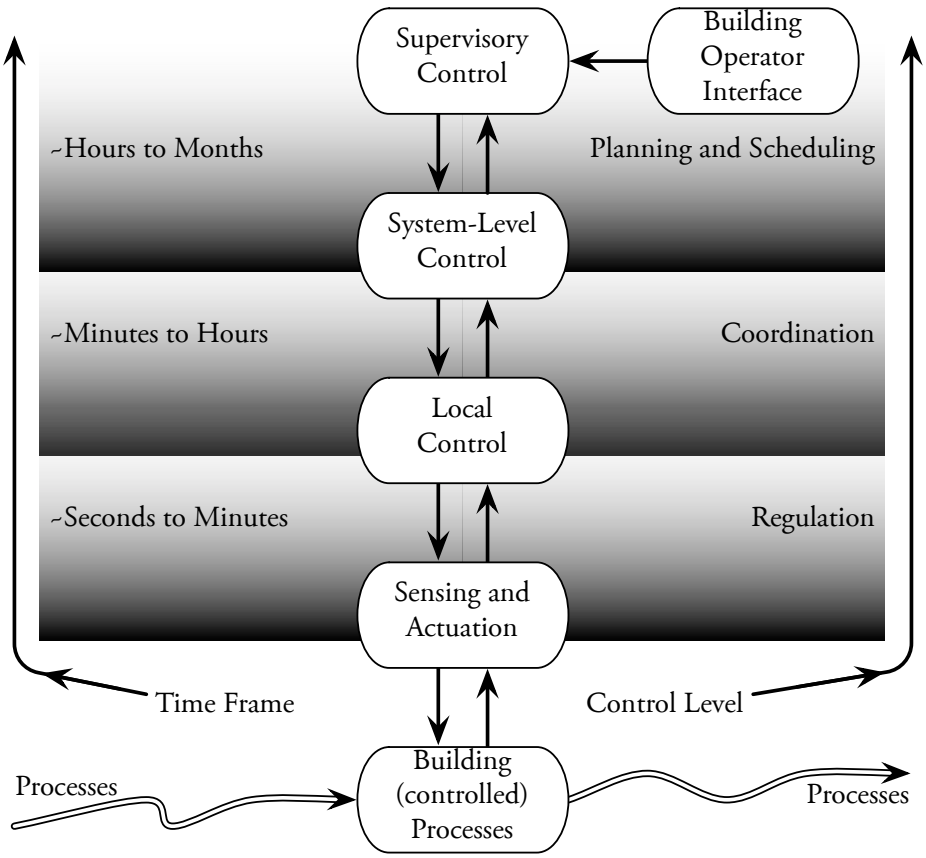

Figure 1.2: Building Automation System (BAS) topology: typical hierarchy for building control.

1.2 Research Motivation

In the development of the modern built environment, engineers have utilized physical modeling and simulation primarily for design purposes. Contemporary methodologies have focused on deterministic approaches for equipment sizing, the relative comparison of design options, and as part of obligations for building (performance) certification programs. Straightforward routines are automated in popular building modeling software; e.g., boiler capacity sized according to the worst-case scenario of maintaining indoor temperature for an unoccupied building on the coldest day of the year. For real-world operations such physical models may be too rigid;[1] i.e., built on limited

[1] The Model-Based Benchmarking (ModBen) project [16] was the impetus for the operational building energy performance research herein; its outcomes are summarized in Appendix A.

supporting data and composed of cascading assumptions, the physical (design) models cannot replicate all dynamical behavior. Although countless development years have gone into building energy modeling tools, there is a dearth of literature or applications showing model predictions matching measured performance in a time-series fashion for more than individual subsystems. There are many studies on calibration, as reviewed in the next chapter, but results can be misleading because model-measurement agreements are only valid for aggregate energy flows. For this reason, building energy model calibration is thought of as "an art and a science." In summary, there exists a great need for thought revision on the suitable mathematical modeling of buildings and related systems, especially in an operational setting for decision support where *uncertainty quantification* is necessary for stochastic optimization of real-world HITL processes.

Combining expert insight, data, and models for decision support is seen as a path forward and has been generically termed *analytics* for its reliance on elements of computer science, statistics, and operations research for business decision making [17]. It is argued that data analytics and physical modeling are central to smart buildings, herein termed an Operational Building Energy Model (OpBEM) and denoting a cyberphysical construct for research purposes. The value proposition is the incorporation of energy performance into, and phrased in terms of, business operations to offer a substantial internal rate of return by complementing and extending the performance gains from energy audits, commissioning, and operations and maintenance (O&M). At the building scale, automated prediction and diagnosis can help focus O&M to save costs while providing superior comfort, leading to increased worker productivity. On a portfolio scale, costly and ill-performing buildings can be identified and systematically mended. Research and development (R&D) into the built environment spans relatively simple devices, such as smart residential thermostats, to the highly faceted smart grid where neighborhoods of commercial buildings and industrial facilities are subsystems. The breadth and depth of these R&D concerns, coupled with ever more demanding expectations, can be addressed from a unified perspective through Bayesian probability.[2]

[2] Dodier [18] identified the various built environment R&D pathways using belief (Bayesian) networks and motivated the Bayesian perspective of later chapters; the fundamentals of Bayesian inference are provided in Appendix B.

For building owners, operators, and other stakeholders, energy as a commodity is obvious at large scales; however, it is more common that energy costs are far outweighed by other business expenditures influencing profit margin; e.g., office worker salaries cost approximately 72 times that spent on energy [19]. Since faulty building performance often stems from practical O&M issues, through a scalable software tool it may be possible to optimally focus manual efforts and thereby save time, effort, and money over a building's lifecycle. Furthermore, smart buildings' participation in the smart grid will require (scenario) forecasting and a portfolio of buildings with flexible loads, whether shaped with the help of energy storage, advanced controls, or both.[3] Although the definition of "building health and energy performance" will vary and is subjective, defining the properties of an OpBEM is a worthwhile pursuit. In this research, focus is placed on describing how buildings operate in the real world by extending popular physical simulations tools, which are modular and allow quick reconfiguration of components, subsystems, and systems, and combining these models with relevant data and analytics so queries can be answered that would be difficult or costly to consider otherwise.

There are numerous modeling strategies for capturing, understanding, and predicting the patterns observed with respect to building energy performance; however, physical design models are often dispensed with upon building construction. Here, it is argued the design model is the "ideal" to be upheld, with its utilization throughout the building's life cycle. Exploring the OpBEM cyberphysical construct is crucial to the built environment for three reasons: 1) buildings are dynamic and some patterns observed in reality are beyond what can be considered with any design model, 2) a given building might be "outside the norm" (e.g., it contains a sizable computer server triggered by worldwide requests) and is thus subject to stochastic concerns outside the realm of traditional, deterministic physical building energy models, and 3) the possible energy and cost savings from an individual building likely do not outweigh the cost associated with sophisticated analytics. OpBEMs are feasible only if automated and scalable.

[3] Renewable energy integration is anticipated to drive power ramping requirements due to solar and wind power generators' influence on supply variability and uncertainty [20]; an examination is provided in Appendix C.

A data-driven toolchain cannot consider all possible situations involving operational building energy performance. It is therefore of value to phrase the problem as one of *process development* and ensure its adaptability, flexibility, and optimization as data are gathered. Inspiration is taken from the application of the Deming cycle[4] to various business problems. Shown in Figure 1.3, the iterative Plan-Do-Study-Act (PDSA) method involves: 1) plan: specific objectives are planned, 2) do: they are implemented, 3) study: they are evaluated to understand how they succeeded and failed, and 4) act: results are acted upon to improve the process in the next iteration. Later chapters argue that the HITL will always be the one to act, but steps 1–3 can be transformed to the building science context as classification, forecasting, and signaling, respectively. The PDSA method has general applicability to the aforementioned OpBEM objectives.

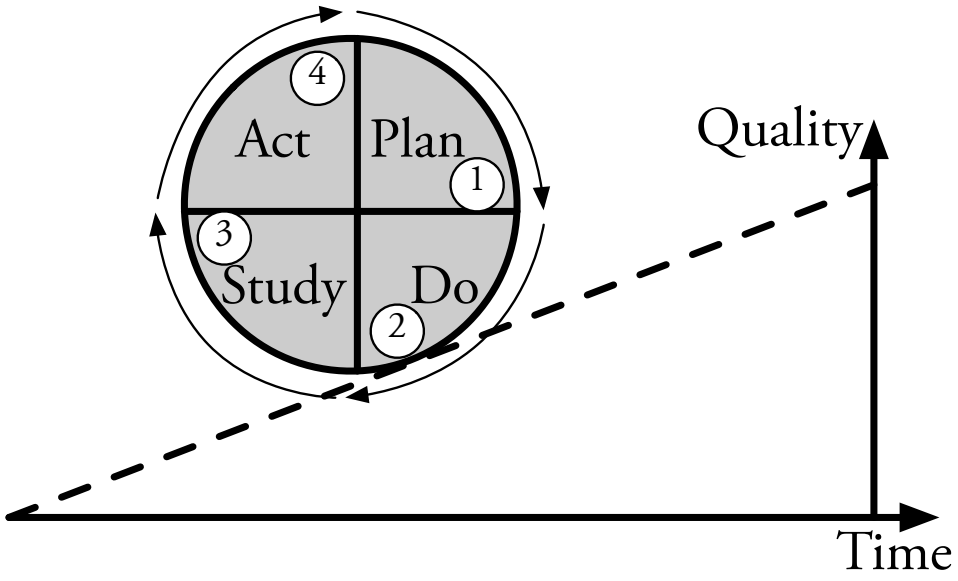

Figure 1.3: The Plan-Do-Study-Act (PDSA) method (or Deming cycle) is an iterative four-step management method used to control and continuously improve processes and products.

[4] William Edwards Deming (1900-1993) was an American engineer, statistician, professor, and management consultant, born in Sioux City, Iowa. He studied electrical engineering at the University of Wyoming (B.S.), University of Colorado (M.S.), and Yale University (Ph.D.). He is most notable known for his work in Japan after WWII, where he adapted statistical process control [21] in a Plan-Do-Study-Act (PDSA) fashion and is credited with propelling Japanese industry from ruins to the second most powerful economy in the world in less than a decade.

1.3 Organization of the Dissertation

The research gap is the substantial differences between the reality of buildings' operational performance and those envisioned by building modeling and simulation studies. The desired outcome is a data-driven toolchain for the operational performance analysis and optimization of buildings. The dissertation is organized into seven chapters, followed by three appendices informing the research perspective taken, as described below.

Chapter 2 *Operational Building Energy Performance* — examines and summarizes publications relevant to the above discussion, as well as tangent applications providing implicit or explicit decision support with predictive or diagnostic considerations.

Chapter 3 *Development of a Data-Driven Toolchain* — a data-driven toolchain for the operational performance analysis and optimization of buildings, expounded in later chapters, is described in terms of an iterative, three-step process utilizing classification, forecasting, and signaling.

Chapter 4 *Classification of Commercial Building Electrical Demand Profiles for Energy Storage Applications* — examines the time-frequency electrical demand profiles of the United States commercial building stock, as modeled from the 2003 Commercial Buildings Energy Consumption Survey (CBECS) data and simulated using the EnergyPlus building energy simulation engine.

Chapter 5 *Comparison of Short-Term Weather Forecasting Models for Model Predictive Control* — evaluates short-term weather forecasting for commercial buildings' model predictive control, with prediction accuracy quantified and compared across various geographic locations using a suite of methods based on traditional time series analyses and artificial neural networks.

Chapter 6 *An Energy Signal Tool for Decision Support in Building Energy Systems* — demonstrates the first prototype of an operational decision support system to aid building operators' insight to ideal and faulted energy performance. A Bayesian approach is used to quantify uncertainty, and the decision of signaling normal or abnormal energy use chosen by minimizing the expected cost of displaying the wrong information.

Chapter 7 *Conclusions and Future Work* — summarizes the results of the research and provides recommendations for its extension, with particular focus on feasible, real-world implementations of the A Posteriori EXplorer (APEX) system concept.

Appendix A *A Summary of the Model-Based Benchmarking (ModBen) Project* — describes project-based outcomes in the development of practical methods and tools to help identify potential building energy savings in a cost effective, timely, and permanent manner; this project was the impetus for the operational building energy performance research.

Appendix B *Fundamentals of Bayesian Inference* — provides supplemental material on the Bayesian perspective to probability and its wide-ranging implications to the research objectives, namely uncertainty quantification.

Appendix C *Identifying Wind and Solar Ramping Events* — describes how wind and solar power can augment uncertainties in the operation of power systems, specifically through increased power ramping variability. Renewable energy integration is playing an increasingly significant role in the electric grid — influencing the need for flexible, smart buildings.

Chapter 2

Operational Building Energy Performance

This chapter reviews the literature pertinent to the stated objective and highlights related concerns, applications, and research desiderata. Perspectives include qualitative and quantitative aspects of operational building energy performance. The goal is to provide context and narrow the scope of research to a thorough, yet manageable examination.

2.1 Buildings in the Digital Age

Building design traditionally starts with architectural planning and programming, moves on to schematic design, follows with design development, and finally, construction documents are produced before the project is bid and built. Of course, as the process develops from conception to completion, there is a decline in the opportunity for modifications which is mirrored by rising costs to change plans. This has been partially combated in modern times through design-build and computer-aided design approaches, resulting in some form of a digital model of the building. The process enhances the speed and accuracy of the building's design and construction, allowing team members to better integrate their ideas and communicate more effectively. The technological movement in building design that began with computer-aided drafting has recently incorporated building information modeling (BIM), which seeks to maintain a digital representation of the physical and the functional elements of the building (e.g., geometry, materials, and specifications) through its construction, operational life, and finally to its demolition [22]. BIM is in its infancy and most focus has been placed on maintaining documentation from conceptional design to as-built construction.

Nevertheless, it has piqued the interest of architects, engineers, and construction personnel alike. Integrated designs will greatly benefit as the technology matures.

Decision support concerning the operational performance of buildings is sparse because the threshold between the proper use and misuse of resources, namely energy, is costly to monitor and magnitudes are subjective; e.g., occupants' influence on electrical demand through plug loads and whether their choices are necessary to operations.[1] It is argued that a data-driven toolchain for the operational performance analysis and optimization of buildings must commence with an examination of what constitutes a 'building model,' as considered in this chapter, and then define the cyberphysical construct necessary for interfacing BIM (or similar) with operational data, simulation of physical models, and HITL processes, as is considered in the next chapter. It is hypothesized the existing set of BAS measurements found in modern, maintained buildings is sufficient for such decision support through an enterprise EMCS; however, the operational data will take uncountably many forms and a process is necessary. It is envisioned the physical simulations used today, currently not well-integrated with the architectural design development, can be rephrased as operational simulation models and build upon historic methods for building simulation and optimization [23].

Operational faults that could be automatically diagnosed are currently left to building operators for manual resolution. The problem is that many faults go unnoticed due to their insidious nature; i.e., major equipment failure is typically reported (by occupants) in a timely fashion but slow moving faults leading to inefficiencies are not. With convergence on a unified, digital building model for the overlapping needs of architecture, engineering, and construction, it is necessary to explore how modeling might be combined with operational measurements. Specifically, a building's life cycle may be extended, costs reduced, or building health enhanced with ongoing energy performance analytics. How operational building energy performance is defined is subjective but closely aligns with building science subfields of fault detection and diagnostics (FDD) and model-predictive

[1] As one might imagine, defining an objective function that involves the value of human actions is controversial and likely impossible in a complex business setting. Thus, human influences on cyberphysical systems are regarded as random processes and research focus is placed on engineered systems associated with (commercial) buildings.

control (MPC) with retrospective and anticipative analyses, respectively. These two research domains make explicit use of building physics models and data; however, abstraction of these tools into processes for decision support is largely overlooked. The following sections summarize research progress and key developments.

2.2 Trends in Building Energy Modeling and Simulation

Authors have detailed the history of building modeling and simulation [23, 24, 25]. It can be summarized to say the sequential developments show its evolution from simple calculations for design purposes to advanced energy simulation. In the mid-1900's, heating was the primary concern of building modeling efforts. A steady-state worst-case heating load was calculated under fixed conditions; e.g., indoor and outdoor temperatures, humidity, etc. Modifications were later made for transmission of solar irradiance and diurnal internal gains. In the early 1960s through the 1980s, analytical methods were developed to treat the lagging effect of thermal capacitance. This was coupled to thorough treatment of geometry and the radiative properties of glazing. Building simulation software was mostly distinguished by its approach to conduction heat transfer through the envelope, with response factor, admittance, and finite-difference methods being the norm. The 1980's saw the development of heating and air-conditioning plant models with simple controls to match energy flows with setpoint analysis; hygrothermal models were then considered in earnest. Steady-state equipment models allowed the comparison of performance for design purposes. Some detailed, dynamic plant models became available; however, with interest placed on longer time horizons for design purposes, these models were usually specifically associated with control investigations. In the 1990s there were a sizable number of commercially available and government-sponsored building simulation codes for use on personal computers. With the wide range of sophisticated system and component models, modelers began to couple computational fluid dynamics (CFD) code to building energy models for realistic ventilation studies requiring conjugate heat transfer at low Reynolds numbers. This "hand-shaking" approach has been considered overkill by some and necessary (modeled) dynamics by others. From 2000 to present, greater modeling detail and new

systems have continually been added to building modeling and simulation programs. To stimulate wider adoption of the software tools, the focus has also been placed on data management protocols and exchange, graphical user interfaces (GUI), and, in general, ease-of-use.

With evermore complex simulation models and designs, building energy modelers are concerned with possible software bugs, deterministic simulation models programmed according to limited data, and the overall need to validate the performance as compared to design [26, 27, 28]. Although the building physics are mostly understood for detailed modeling and simulation, there are considerable uncertainties associated with the use and performance of real buildings [29]. These include, but are not limited to, occupancy and behavior, micro-climate effects, and significant deviations from the assumptions used to build the model. Of course, the overwhelming majority of building energy models are deterministic even though stochastic considerations are more fitting in light of unpredictable human influence. Building engineers are looking to reevaluate existing approaches with measured data toward the enhancement of whole building energy performance. A few viable ideas include:

(1) Maintain a database of measured building performance [30] and notable events under a given supervisory control strategy and ambient or internal influences, e.g. typical weather, known equipment malfunction, etc.

(2) Have systems in place to easily allow occupants to deliver feedback [31] to performance stakeholders (e.g. building operator, manager, owner, etc.), as well as to quantify the occupants' impact on the building for bilateral feedback.

(3) Adapt the supervisory control of the building in real time, and also forecast future events or probable time series trajectories, according to weather and the needs of the electric grid or other enveloping systems [32].

2.2.1 Sensitivity Analysis

Monte Carlo approaches are the most commonly used methods to understand numerical sensitivities [33]. Reddy and Maor [34] used a midpoint Latin Hypercube Monte Carlo (LHMC) stratified sampling on three levels. The steps are to: 1.) assign probability distributions to each building model inputs or parameters, in particular, a triangular distribution, 2.) the triangular distribution is discretized into low, medium, and high according to three ranges leading to equal probability, and 3.) randomly select LHMC combinations (high/medium/low) for the inputs and parameters. Each LHMC parameter combination becomes a realization for a building energy simulation. Upon completion of the Monte Carlo simulation, measured data is compared to the simulated results. Those with close agreement are included in the calculation of a chi-squared statistic. If the samples used in the chi-squared calculation are non-uniform, the more sensitive the measured and modeled data agreement is to a particular high/medium/low combination. An optimization can focus on adjustments to these influential inputs and parameters.

2.3 Data-Driven Assessments

Dodier et al. [35, 36, 18] researched prediction and diagnosis in engineering systems with belief (Bayesian) networks, forming the basis for research presented herein because of the uncertainty quantification aspects. Kaldorf et al. [37] detailed their practical experiences from developing and implementing an expert system diagnostic tool, with FDD for underperformance or abnormalities. The faults considered being total or partial component failure, wrong parameter settings, operator errors, undersized system capacity, changes of building, zone usage, etc. Song et al. [38] examined the online simulation of whole building energy consumption for the concerns of fault detection and optimization. It was accomplished through the development of a heating, ventilation, and air conditioning (HVAC) automation system, embedded in the energy management and control system.

2.3.1 Typical Building Energy Performance Data

Before combining energy models with measured data, it is necessary to understand what data are typically available. In general, from coarse to fine resolution, typical data include the below descriptions.

Utility Billing Data Are a record of energy performance to an extent [39]. Regression models can be used to compare and contrast similar buildings, validate pre- and post-retrofit measures and, of course, the calibration of building energy simulation models according to aggregate energy flows.

Interval Data Take utility metered, whole-building data to the next level of resolution [40] with demand or consumption provided at time intervals typically ranging from 15 minutes to 1 hour. Comparing intervals of sequential days over a time period allows building energy patterns to emerge. This is helpful for troubleshooting issues associated with whole building energy performance such as the poor scheduling of equipment; e.g., fully lighting an unoccupied building at night.

Submetered Data Are specific to a particular building or building system [41], with the time sampling and spatial distribution of the measurements specified according to the needs of a given project. Typically efforts are made to decompose the energy performance of the building into end-uses such as heating, cooling, lighting, etc.

Automation System Data Are obtained from the BAS or EMCS [42]. Nearly all control data can be obtained from local and system controllers. Data are sampled, stored, and trended according to user queries. Alarms can be activated according to user-programmed logic or conditions. Acquiring data is constrained by the bandwidth of the network, the speed and storage limitations of the acquisition system.

An ongoing measurement campaign may harness all data available and require additional measurements. It may only require interval data. The answer depends on the application and the utility of data with regard to decision making. Frequent and spatially diverse data are helpful for informing a detailed physics model of the building, at least when assessing whole-building energy

performance. Simple time series models using interval data may be adequate for assessing simple energy performance (e.g. cost per square foot) or forecasting weather, thermal load, and electrical demand over some operational horizon. A range of applications exist according to the data available and model employed.

2.3.2 Calibration, System Identification, and Evidence-Based Approaches

There may be tens to thousands of inputs and parameters, each with considerable levels of uncertainty, required to calibrate a building energy model. Even so, there will never be a perfect agreement between model output and measured data. It is most common that the problem posed is one of optimization, with the objective function using various metrics or distance measures to quantify the disagreement between model and reality. When the model is *calibrated* it is typically up to the user, and that is why some refer to calibration as more of an art than a science. It is a highly underdetermined problem, and a multitude of solutions may satisfy a minimum according to the objective function criteria, and that is why so much interest has been put into its solution. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14 [43] standardizes a set of energy, demand, and water savings calculation procedures to provide guidance on minimum acceptable levels of performance in determining savings; normalized mean bias error (nMBE) and coefficient of variation of the root mean square error (CV-RMSE) metrics are utilized.

Reddy and Maor [34] give a comprehensive account of building energy model calibration, with the identification and summarization of the best techniques and approaches in current practice. A methodology of parameter estimation and systematic calibration is developed for assessing and quantifying uncertainty. A four-step methodology was proposed: 1.) identify the most influential parameters through an audit or heuristics, 2.) perform a Monte Carlo simulation for a course calibration leading to feasible solutions, 3.) use a fine grid on the feasible solution space to manually calibrated or via optimization, and 4.) compute the calibrated models prediction error for uncertainty quantification.

Evidence-based approaches [44, 45] depend on prior knowledge through blueprints, and other reliable sources of building information, so that results can be reliably reproduced. This approach recognizes that adjustments based on poor sources will inevitably lead to problems, and a systematic approach with concerns for sensitivity is necessary. Raftery et al. [45] proposed three elements crucial to the calibration: 1.) detailed models that represent the real building accurately, 2.) methodology reproducibility, and 3.) abundant measured data. Norford et al. [46] suggested a two-stage manual process: 1.) tune tenant-based inputs and HVAC schedules, and then 2.) tune HVAC equipment and building envelope parameters. This was based on occupant behavior having the most significant impact on energy consumption. Lee and Claridge [47] investigated an automated and iterative calibration routine as part of a building case study, claiming the routine's ability to filter local minima was its advantage over manual calibration. Other authors have noted the helpfulness of interval data due to the ability to discern occupant patterns [48], and the ability of high-frequency submetered data to complement utility data [49].

Since building energy simulation models typically rely on a small subset of the total number of inputs and parameters for their calibration, focus on a small number of crucial factors can save calibration time. Although this can be accomplished using expert experience, it often makes sense to perform uncertainty and sensitivity studies to understand what impacts the model the most. Eisenhower et al. [50, 51] studied the influence of approximately 1000 parameters on complex building energy models, with focus on efficient uncertainty and sensitivity decomposition, allowing accurate meta-model fits from a handful of parameters and an understanding of which parameters' sensitivities are best for optimization. Burhenne et al. [52] used a Monte Carlo based methodology for uncertainty quantification, using building simulation coupled to cost-benefit analysis, to determine model inputs that lead to positive net present value investments.

2.3.2.1 System Identification

Building energy modeling and measurement are typically associated with new construction. However, 93% of the United States commercial building stock was built before 2003 and 74% built

before 1989 [53]. For the residential stock, 70% of homes were built before 1989 and new construction annually accounts for approximately 1% of the building stock [54]. These statistics highlight the impact of existing construction. To truly address energy availability, security, and efficiency, building modeling and measurement for the enhanced performance of existing construction is a logical first step. Furthermore, combining insights and lessons learned to new construction will lead to a shift in the energy performance of the building stock over time. There is a great need for accurate building energy models combined with measured data. Movement away from theoretical simulated performance to actual delivered, data-driven performance is necessary to meet energy goals, codes, and owner requirements.

System identification is used as part of analysis, simulation, prediction, control, optimization, and fault detection [55, 56, 57]. ASHRAE distinguishes between a forward (classical) approach and a data-driven (inverse) approach [58]. The former is mere simulation, and the latter constitutes various forms of system identification. First principle models are extremely helpful for simulation and quantifying expectations for design purposes, but there is often a tremendous amount of assumptions that go into their development. For design purposes, one can always make conservative assumptions and overengineer a given system. For data-driven models, design assumptions are replaced by uncertainties in the mathematical description of the model; e.g., parameters, inputs, structure, etc. The application of the model and data available will likely drive the system identification methodology.

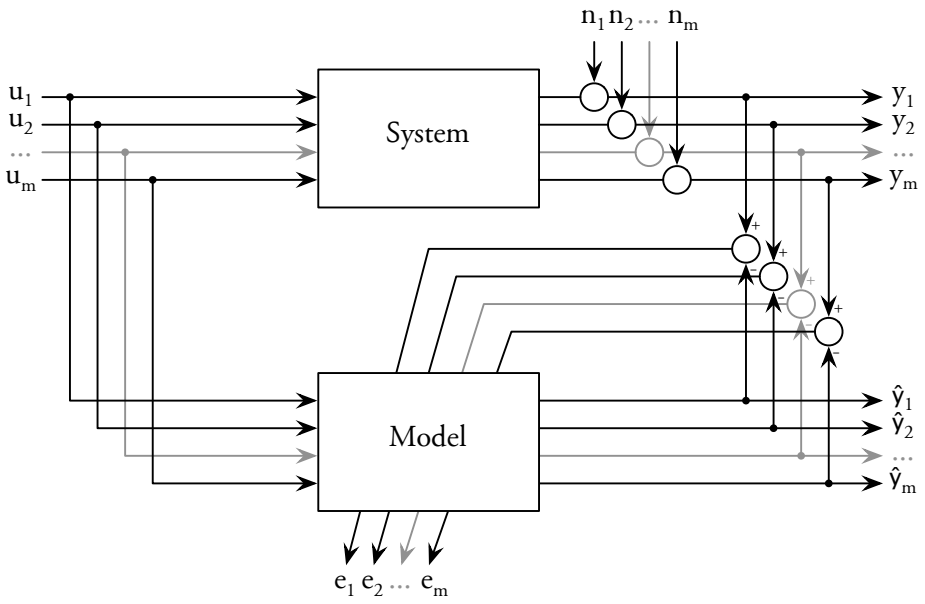

Figure 2.1: Typical multiple-input, multiple-output (MIMO) system identification. Various algorithms are used to tune the model according to its prediction error.

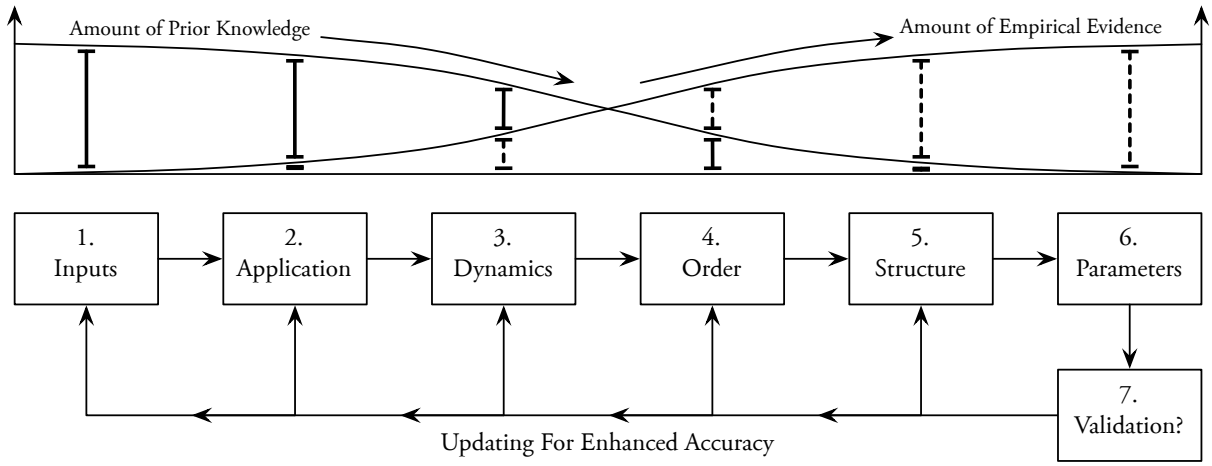

Figure 2.2: The system identification procedure can be roughly broken into six stages, with the final, seventh validation stage used for accepting or revising the model as necessary; adapted from Nelles (2000) and modified according to a Bayesian perspective.

Figure 2.1 shows the system identification problem to be addressed. The problem can range from simple single-input, single-output (SISO) identification to complex multiple-input, multiple-output (MIMO) identification. In most realistic problems there is a limited number of observations available pertaining to inputs, outputs, variables, and parameters. Prediction error over a given time period, combined with uncertainty in the measurement of system output, allows the model to be updated using various algorithms to obtain superior modeling accuracy. The model itself can range from simple linear time series models to detailed physics models with stochastic considerations. Using a deterministic, first principle building energy model is possible, often termed *calibration*, by using all information available (e.g., blueprints) combined with reasonable assumptions for missing values unknowns. Since the problem is highly underdetermined, there is an uncountable number of variables or parameters that could lead to the same observed output, and the model's accuracy is a function of the quality of information. It is of interest to quantify these phenomena and provide a systematic means for evaluating a given model's appropriateness according to the problem being addressed.

Figure 2.2 shows the primary elements required for system identification [55, 56, 57]. During the model development, prior knowledge (i.e., understanding of the physics or data from previous experiments) is large at the start of the system identification and decreases with time. Mirroring the use of prior knowledge, the amount of empirical evidence is small at the start of the model development and increases with time; i.e., the prior to posterior evidence. Note that a temporal aspect to the procedure is not explicitly included in the figure, and the validation stage may require movement back to a previous stage.

With an emphasis on applications associated with building energy models, the required seven steps for system identification [56] can be summarized as follows. Since all modeling requires unique considerations, the general approach described below is not exhaustive.

(1) Inputs — the physical processes influencing a given variable are usually known to some extent. However, some inputs, such as occupant influences, are hard to quantify due to

their stochastic nature. Some inputs may be recorded, like building temperature setpoints, while other inputs are known to occur but not directly measured. It is an intractable problem to measure all building inputs to establish a model, and some combination of sensitive inputs must be chosen. The crucial inputs necessary can be identified according to expert knowledge or through (machine) learning, including unsupervised or supervised approaches.

(2) Application — involving the highest level of engineering expertise, incorporating prior knowledge toward satisfying the nature of the problem is difficult. The system identification might have applications of simulation, classification, diagnostics, etc. Computational requirements may influence the dimensionality of the problem considered, and offline or online model usage drives the adaptability and complexity of the model.

(3) Dynamics — the resolution of the model's dynamical representation will drive the variable relationships that must be explicitly considered. For instance, the electrical demand of a given building may be the primary focus, and detailed air distribution system modeling is required due to very large fans. Or frequency regulation for grid ancillary services [59] may require detailed fan models only; other variables can be lumped together or parameterized according to the application of system identification.

(4) Order — increasing the complexity and dimensionality of a model can lead to higher fidelity to measured performance, but trade-offs exist between dynamic fidelity and static approximation accuracy. A higher-order and lower-model may both be inaccurate in different senses of bias and variance, but there exists a "correct-order" model according to the data and the dynamics explicitly modeled.

(5) Structure — if a purely physical model is considered, the model's structure is fundamentally tied to earlier steps in the system identification procedure. However, the model may be a hybrid of physical and statistical relationships. For example, an industrial building may

have a process that consumes 90% of all energy at the facility. A detailed physics model of that process may exist, but may only weakly depend on surrounding processes which can be approximated by simple statistical relationships. An appropriate structure emerges according to the utility of various subsystems.

(6) Parameters — a given subsystem may consist of numerous variables that can be parameterized in various fashions while equivalent model accuracy is maintained. The choice of parameterization may or may not be important. This is a function of the data that are available and how the model will be used. If a human must interpret the parameter value for decision making, it is helpful if the parameter has a physical interpretation. Furthermore, fiddle parameters may be necessary as part of the model development; e.g., sensitivity to initialization values.

(7) Validation — the model quality must be evaluated in light of its ultimate goal. Unfortunately, this step is iterative and highly dependent on the problem domain. The utility of a given subsystem may inform how accurate its model need be; however, errors can propagate throughout the system as a whole, and the aggregate error is hard to attribute to a given subsystem. A methodology for model validation must be resolved for the application of system identification and be specific to the domain of interest.

2.3.2.2 Clustering of Energy Performance

Regression models of steady-state energy performance correlate energy consumption with average temperatures or degree-days, typically using one to five parameters [58]. They are quite helpful for validating pre- and post-retrofit energy conservation measures (ECM) [60]. Their use has been criticized for not being able to simultaneously handle both heating and cooling aspects [61] and poor performance in the estimation of building parameters [62], with more sophisticated dynamic models suggested. Nevertheless, for their intended use, successes range from the inclusion and prediction of non-weather-related variables [63] to HVAC operating mode [64] using occupancy

rate as a proxy. The effect of time resolution on multiple linear regression was investigated by Katipamula et al. [65], with results indicating daily time scales are most beneficial for retrofit savings determination, and hour-of-day models are best for O&M. The uncertainty of measured energy savings has been investigated [66], but, in general, uncertainty quantification of steady-state energy performance in the literature is sparse.

Numerous researchers have tried to classify patterns influencing building energy performance as a function of weather or other variables. Katipamula and Haberl [67] developed a methodology for day-type identification using monitored non-weather dependent end-use data, allowing typical load-shape data to be generated. Hadley [68] identified typical weather day types using National Weather Service data with principal component and cluster analyses. It was possible to group HVAC system energy consumption into a few typical days. Seem [69] identified day-types of similar energy consumption profiles using a pattern recognition algorithm. Data features were extracted from the time series of energy use and grouped into seven clusters after an outlier detection scheme was used to aid grouping. Applications of supervisory control and abnormal energy consumption were noted. The research was extended [70] with field tests in buildings having: 1.) poor chiller control strategy and subsequent failure, 2.) poorly designed HVAC equipment, and 3.) following an electrical panel change, improper operation of equipment. More recently, Miller et al. [71] have used a day-typing process based on Symbolic Aggregate approXimation (SAX), motif and discord extraction, and clustering to detect the underlying structure of building performance data. Discords, or infrequent daily patterns, are filtered and tagged for future analysis; Motifs, or the most frequent patterns, are detected and further aggregated using k-means clustering. Results are correlated with normal and abnormal energy consumption day-types.

It is apparent as researchers strive to assess energy performance on shorter time scales, and spatially as well, quantification of energy performance maps to whole-building fault detection – where a fault is any abnormal deviation from expected behavior. Of course, in practical applications, there are limits on data and the accuracy of physical models. Coupling measured data to simulations helps impute but comes at the cost of considerable assumptions.

2.4 Fault Detection and Diagnostics

FDD has mostly been associated with critical processes; e.g., nuclear technology, aeronautics, and the chemical industry. Pioneering research in the automotive industry is attributed to Isermann [72] and mirrors the noncritical process concerns of building operational performance analysis and optimization.

2.4.1 Foundational Applications in Buildings

Katipamula and Brambley [6, 5] provided an exhaustive survey of FDD literature as applied to buildings citing "poorly maintained, degraded, and improperly controlled equipment wastes an estimated 15% to 30% of the energy used in commercial buildings." In the first article, a discussion is provided on FDD, embodied by three key processes, fault detection, fault isolation, and fault identification, and how these topics have been used in the aerospace, automotive, and chemical fields, but only limited building applications. They classify diagnostic methods into 1.) quantitative, 2.) qualitative, and 3.) process history approaches. There are extensive sources cited according to their classification scheme, and the reader is referred to the treatise for the breadth and depth of FDD concerns. The second article focuses on FDD in buildings, separated into refrigerators, air conditioners and heat pumps, chillers, and air-handling units; statistics and prognostics on failed building equipment are not considered. The current state of diagnostics in buildings is discussed along with comments on future applications. Typical building energy faults are described in Hyvarinen et al. [73] as part of an expert survey from the International Energy Agency (IEA) and, combined with the author's personal experience, are condensed in the following tables to illustrate the vast number of faults encountered in building operations.

2.4.2 Building Primary and Secondary Systems

Over 30 years ago, HVAC system fault detection and diagnosis have been studied [74] and suggestions made concerning the fact that energy management and control systems could be used

Table 2.1: Common faults found in chillers.

Refrigerant-based Faults	**Pressure-based Faults**
lack of refrigerant	loss of vacuum
presence of air in the refrigerating circuit	clogging of condenser tubes
presence of refrigerant in the lubricating oil	clogging of evaporator tubes

to detect and compensate for failures in HVAC equipment. The field has seen numerous studies seeking to explore this area in greater depth.

2.4.2.1 Chillers

Braun [75] reviewed research related to automated fault detection and diagnosis for chillers, packaged air conditioners, and other vapor compression cooling equipment, assessing the state-of-the-art in FDD for said equipment. Haves and Khalsa [76] provided a similar review. Yu and Chan [77] investigated the energy signatures of chillers (e.g., kWh per unit floor area of a building in m^2) under various design options and operating strategies. Andersen and Reddy [78] provide an Error In Variables (EIV) regression approach for chiller performance data. Wang [79] developed dynamic models of a centrifugal chiller for online control strategies. Wang and Cui [80] used principal component analysis to capture the correlations among variables in centrifugal chillers, allowing robust fault detection and diagnosis based on six performance indices describing the health condition of centrifugal chillers. Other authors have considered parameter-based approaches for FDD [81], forecasting [82], model-based assessment [83], rooftop units [84], fault isolation [85], and sensor faults [86, 87]. Table 2.1 provides the most important faults common to chillers.

2.4.2.2 Boilers and Heating Systems

Huang et al. [88] developed a real-time fault detection and diagnosis system for control of thermal power plants in large-scale industrial applications. However, HVAC field studies of FDD for heating systems and boilers are sparse. Table 2.2 shows the most important faults common to heating systems and boilers.

Table 2.2: Heating system and boiler faults common to practice.

Design-based faults	**Pressure-based faults**	**Other faults**
zone temperature too low	defects of manometers (boiler, gas expansion system)	bad boiler efficiency
boiler, radiator or pump size wrong	leak of the gas tank	defects of sensors (water flow, gas flow)
defects in valves (three-way valve, non return valve)	blockage of boiler pipes and heat exchangers	zone temperature too low
	leaks in pipes	thermostatic valve defect
	defects of manometers (boiler, expansion system)	

2.4.2.3 Air Handling Units

Lee et al. [89, 90, 91] used artificial neural networks for fault diagnosis of air-handling units (AHU) by uncovering the steady-state relationship between the dominant symptoms and the faults, as well as for temperature sensor recovery. The research was extended [92], but instead used residual and recursive parameter identification methods (e.g., ARX and ARMAX) for both SISO and MISO. Karki and Karjalainen [93] developed AHU performance factors for monitoring and defining performance criteria. Chen and Braun [94] implemented rule-based FDD for packaged air conditioners, having low hardware and software requirements, and showed good performance in laboratory experiments. Schein and Bushby [95] also developed rule-based approaches for AHUs and variable air volume (VAV) boxes with performance assessment rules and control charts. Other authors have considered functional testing [96], lifecycle performance [97], economizers [98], calibration [99], sensors [100], and BAS-based approaches [101].

2.4.2.4 Air Mixing Section

Table 2.3 shows the most important faults common to an AHU's air mixing section. It is noted that humidity sensors are not very common in practice, but, when used, relative humidity sensors are most common in the return air ducts; furthermore, measurements on outdoor air conditions are more common than on mixed-air conditions.

Table 2.3: Common faults in AHU's air mixing section.

Sensor faults; i.e., temperature or humidity (dew point) on return, outdoor, and mixed air	**Airflow faults**; i.e., damper and actuator (relief, recirculating, and outdoor air)	**Controller faults**; i.e., mixed air components of AHU
complete failure	stuck (open, closed, intermediate position)	control signal (no signal, incorrect signal)
incorrect reading (offset, wrong scale, drifting)	incorrect minimum positioning of outdoor air damper	software error
excessive noise	air leakage past damper when closed	improper tuning (unstable, sluggish)
	faulty indication of damper position	

Table 2.4: AHU's filter-coil section common faults: Part 1 of 3.

Temp. sensor faults; e.g., supply air, water entering or leaving preheating coil, water entering or leaving cooling coil, freeze protection)	**Humidity sensor faults**; e.g., supply air	**Filter faults**
complete failure	complete failure	partially clogged
incorrect reading (offset, wrong scale, drifting)	incorrect reading (offset, wrong scale, drifting)	incorrect/malfunctioning DP sensor signal
excessive noise	excessive noise	leakage through or around

2.4.2.5 Filter-Coil Section

Tables 2.4 through 2.6 show the most important faults common to an AHU's filter-coil section. It is noted that 1) many functions of the filter section are built-in to a unit, originally set to a value by the manufacturer and never touched again, and 2) the performance state of the filter is manually reset when filters are changed. Custom units involving one or more pressure sensors, which can be read by the BAS, are not usually calibrated correctly and have to be set by the technician. Furthermore, feedback is unusual on damper positions; packaged units may include this function, but usually, the command signal is assumed to be correct. While most plumbing tends to be installed correctly, mechanical contractors sometimes try to hide mistakes or assume the controls person can work around 'as-built' plumbing modifications made during install.

2.4.2.6 Fan Section

Table 2.7 provides the most important faults common to an AHU's fan section. Contemporary practice is to run a variable-frequency drives (VFD), which of course need to be configured correctly. Packaged units are usually set up correctly but there are fewer control options. Some units have VFDs that need to be manually set up, causing a litany of possible faults. It is noted that pressure sensors are influenced by the building's design. For instance, as the tube's length increases so do the probability it is pinched or damaged – leading to serious issues.

Table 2.5: AHU's filter-coil section common faults: Part 2 of 3.

Valve and actuator faults; e.g., preheat coil or cooling coil valve	**Coil faults**; e.g., preheat or cooling coil)	**Supply air temperature (components of AHU) controller faults**
stuck (mechanical failure, actuator/motor failure)	fouled coil	control signal (no signal, incorrect signal)
water leakage past closed valve	partially plugged coil	improper sequencing of valves and dampers
faulty indicator of valve position	wrong coil installed (oversized, undersized)	software error
improper installation (installed backwards, mixing & diverter ports interchanged)	water leaks	poor tuning (unstable, sluggish)
wrong valve installed		
poor valve authority		
clogged valve		

Table 2.6: AHU's filter-coil section common faults: Part 3 of 3.

Plumbing equipment faults	**Plumbing system faults**	**Other plumbing faults**
pump failure (3-way preheating/cooling coil valve applications)	complete failure	cavitation
partially clogged	poor efficiency	wrong pump installed
piping partially blocked	silted up	water pressure
too high/low pressure in hot/chilled water supply line	water temperature	too high/low hot water supply temperature
too high/low hot water supply temperature	too high/low chilled water supply temperature	water leaks

Table 2.7: Faults common to an AHU's fan section.

Fan; e.g., supply and return fan	**Pressure sensor**; e.g., supply duct	**Flow measurement station** e.g., supply and return duct	**Fan controller component of AHU**
complete failure	complete failure	complete failure of sensor signal	control signal to fans (no signal-correct signal)
stuck (full speed, intermediate speed)	incorrect reading (offset, wrong scale, drifting)	incorrect reading (offset, wrong scale, drifting)	improper flow rate differential set point
inlet/outlet vanes (stuck, failed actuator)	excessive noise	excessive noise	improper pressure set point in supply duct
wrong fan installed	improper location		variable speed drive malfunction
	poor resolution/accuracy		software error
	deterioration of sensor with time		poor tuning (unstable, sluggish)

2.4.2.7 Building Secondary Systems

House et al. [102] considered air distribution from the perspective of controls and diagnostics, with a method that enables operational characteristics of individual HVAC components to be extracted from high-frequency measurements of whole-building power. Schein and Bushby [103] approached FDD from the system level with a hierarchical rule-based fault detection and diagnostic method, allowing a human operator to interface with multiple, equipment-specific FDD tools.

2.4.2.8 Variable-Air-Volume Boxes

Seem et al. [104] presented fault detection of VAV boxes in laboratory experiments and from real building data, using an indices-matching algorithm to isolate faulty actuators and valves. Schroeder and Bradford [105] presented results from a model-independent FDD approach to VAV boxes, using performance indices as well. Table 2.8 and 2.9 shows the most important faults common to VAVs. It is noted that VAV controllers have tubing from the controller to the airflow ring. As such, tubing caps (i.e., pressure measurement points) can break or plugged incorrectly – leading to poor damper control due to damaged actuation. Furthermore, shafts can break or come loose from the attachment, leading to failed control. Although feedback control is used in VAVs, damper position is only checked at the time of install and can lead to problems if set incorrectly.

2.4.3 HVAC Control Systems

Fasolo and Seborg [106, 107] used statistical quality control charts for fault detection in HVAC control systems, and monitored performance to determine a controller's performance index. Salsbury and Diamond [108] used model-based feedforward control for fault detection in HVAC systems, with results from testing the controller with a simulated dual-duct air-handling unit. Hao et al. [109] considered an HVAC monitoring system for fault-tolerant control and data recovery, using principal component analysis to distinguish between normal and faulty behavior. Other authors have focused exclusively on sensors and static nonlinearities [110] as part of FDD.

Table 2.8: Common VAV faults: Part 1 of 2.

Damper and actuator	**Reheat coil**	**Plumbing**	**Reheat valve and actuator**
stuck (open, closed, intermediate position)	fouled coil	piping	stuck (mechanical failure, actuator/motor failure)
air leakage past closed damper	partially plugged coil	partially blocked	water leakage past closed valve
incorrect minimum position	wrong coil installed	silted up	faulty indicator of valve position
faulty indicator of damper position	water leaks	water/pressure (high/low)	installed backwards
		water temperature (high/low)	poor valve authority
		water leaks	plugged valve

Table 2.9: Common VAV faults: Part 2 of 2.

Flow measurement station	**Zone temperature sensor**	**VAV box controller**
complete failure of sensor signal	complete failure	control signal (no signal, incorrect signal)
incorrect reading (offset, wrong scale, drifting)	incorrect reading (offset, wrong scale, drifting)	improper sequencing of valve and damper
excessive noise	excessive noise	software error
		poor tuning (unstable, over controlled, sluggish)

Chapter 3

Development of a Data-Driven Toolchain

This chapter explores the research question in greater detail, identifies the theory necessary to develop the engineered system, and argues that practical progress toward a functioning OpBEM can be made using a three-step process. Specifically, an iterative, three-step process is developed to 1) classify building energy performance scenarios, 2) forecast dynamics over a planning horizon of interest, and 3) signal human decision-makers concerning deviations from ideal behavior.

3.1 Research Statement

Buildings reside at an important juncture between humans and the engineered systems that support their existence. Prefixing *smart* to buildings is argued to be most important when both individual and societal impacts can be changed in a positive manner with seamless transition from old technologies; e.g., continued progress in obtaining energy from secure, reliable, and renewable sources, while simultaneously using it in an efficient manner. *Smart buildings* are a future resource for cyberphysical optimization, whether at the building or portfolio level, and are an important HITL system in which to consider health. As such, 'building health' could indicate both quantitative, building-like-a-machine aspects and qualitative, human-interpretable aspects. The former could be daily kilowatt-hours consumption, and the latter some form of "today's building energy performance was rated as green," where metrics map to the color to signify 'good' or 'sustainable' and opinions (e.g., comfort complaints) can be recorded jointly for analysis. Ultimately, cyberphysical optimization problems will involve many or all business processes.

Appreciable levels of uncertainty and subjectivity will prevail in the operational performance analysis and optimization of buildings. Uncertainty is inherent to these mathematical models, not only due to unobserved dynamics but from occupants' stochastic behavior and preferences, weather forecasts and microclimate, as well as uncountable other factors. Subjectivity is embedded whether considering a building's impact on humans or, conversely, humans' impact on buildings. Automating the discovery of an OpBEM involves the mathematical modeling of whole-building dynamics, including all system and subsystem information necessary for a smart building, while considering humans' input and objectives. It is posited that the next level of automation required of smart buildings is probably not yet full artificial intelligence (AI), which may not be desirable due to concerns of the proverbial super-intelligent AI takeover [111], but rather can be expected to rest on a probabilistic expert system foundation mixed with HITL considerations. Taken a step further, smart buildings, and the inherent machine-to-machine communications necessary for smart grid applications, are only as effective as their design for humans and operation by humans. In broader terms, cyberphysical designs must support not subvert nor supplant humans with AI — a sort of golden rule of the information age — and the built environment is a natural place to define human-machine interfaces (HMI) for the better.

Smart buildings can be designed or retrofit to guide building performance under changing objectives and conditions, with this decision support facilitated by supervisory control and expert fault intervention (i.e., if needed). However, a distributed approach is necessary for buildings and will likely built on top of an EMCS. In short, smart buildings will have to co-evolve with the smart grid and the obvious driver for change is renewable energy integration. Power systems planning has historically been addressed from a centralized perspective where small counts of large savings were sought. Energy systems optimization is typically fashioned according to bottom-up planning. Strategies for each must be better aligned if progress toward any smart system is to be made. Toward the goal of enhancing building performance with a magnitude relevant to the smart grid, applications may be managed through software-as-a-service (SaaS) hosted in the cloud.[1]

[1] This research seeks to produce a data-driven toolchain, for the operational performance analysis and optimization

3.2 APEX Conceptualization

The A Posteriori EXplorer (APEX) system concept is shown in Figure 3.1; it is considered to be a concept because research herein focuses on the iterative, three-step process it takes to train the underlying algorithms and framework. In essence, APEX's allows buildings operators to guide the Bayesian learning approaches of Appendix B with an OpBEM template of their choosing. It helps reduce uncertainties in building operations, and thus enhances building health and energy performance. The tool, like any tool, cannot automatically function at high levels without human guidance. Experts will always have to be the final decision makers; however, APEX seeks uncertainty reduction is a top-down fashion, as shown in Figure 3.1, to replicate the human learning process (left-to-right):

$$data \rightarrow information \rightarrow knowledge \rightarrow insight \rightarrow wisdom$$

It is envisioned that measured data, analytics from the simulation of physical models, and expert recommendations will be used to further define the OpBEM. The design of a template system will allow replication and aggregation of trained OpBEMs. Automating these evaluations and advising their enhancement through model-based, whole-building prediction and diagnosis is APEX's function; the below subsections examine its qualitative aspects.

3.2.1 OpBEMs for Smart Buildings

Stakeholders interested in the performance of buildings are often dismayed at gaps between expectations and reality. Although design may have the largest impact over a building's lifetime, building health and energy performance is rarely tracked to ensure design goals. OpBEMs could be a valuable tool to automate and enhance human decision-making processes, largely missing in practice. Uncertainty and subjectivity act to filter conclusions drawn, muddying both the precision and accuracy of such tools, and perhaps the difficulties or complexities have dissuaded their invention. However, the time humans spend indoors, with their direction over controls and subject

of buildings, through a three-stage, interative process of classification, forecasting, and signaling.

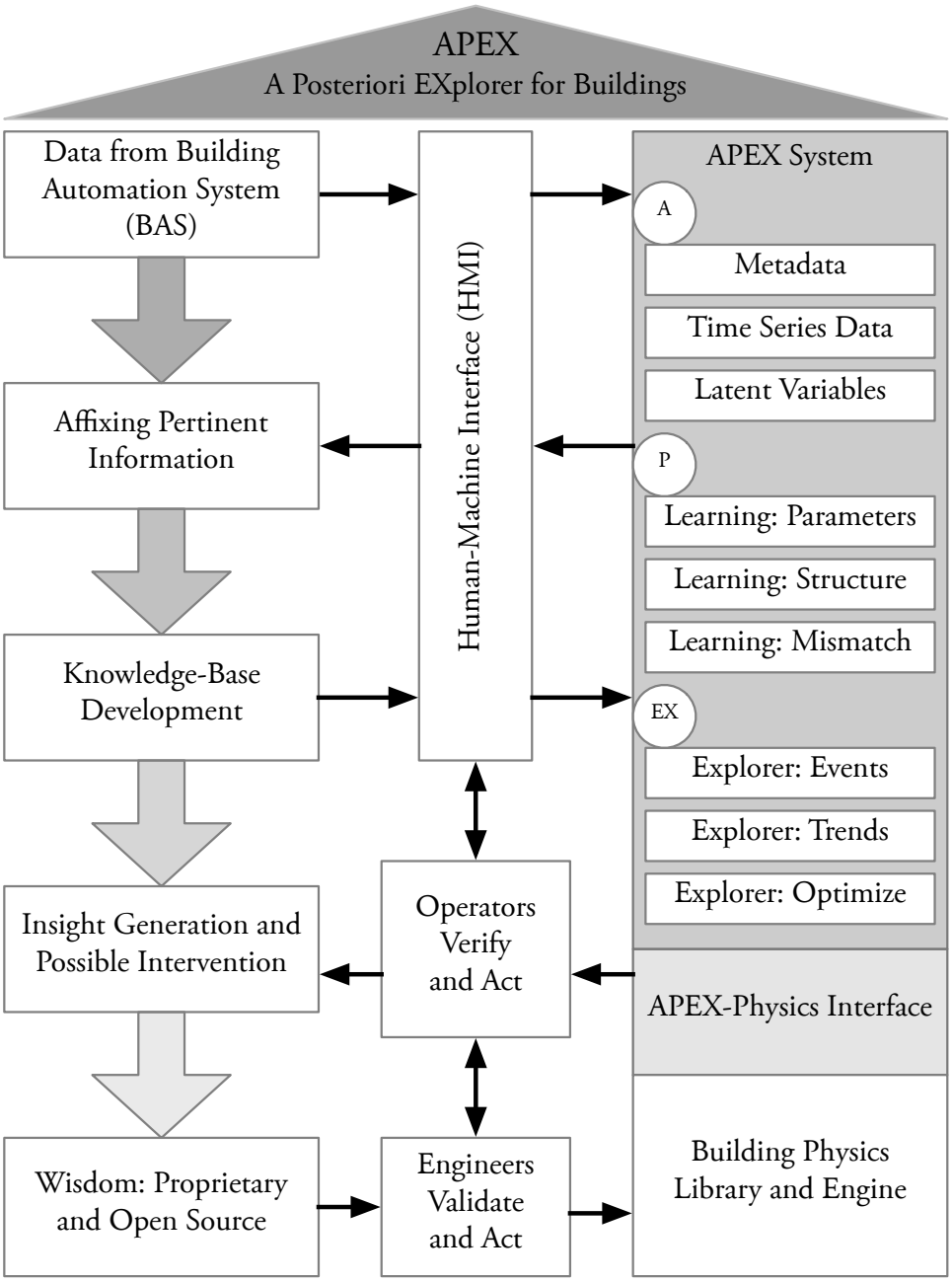

Figure 3.1: A Posteriori EXplorer (APEX) system concept for buildings.

to the control environment, incentivizes the development of OpBEMs

The ability to act on predictions and diagnoses is of great concern to optimizing a building's operational energy performance. Fundamental to decision support is the subjective assessment of building health and performance at any moment and effective planning over various time horizons of interest. It is posited that 1) causal models are most helpful because they track human thought processes, 2) humans should always be in the loop in terms of both occupant satisfaction and top-down decision making, and 3) automating predictions and fault diagnostics in buildings is tantamount to smart buildings. Thus, automation to direct humans' manual effort, concerning actions they might take, begets manual diagnostics and acts as a force multiplier of nominal O&M. It also enhances supervisory control, for instance, with advanced topics (e.g., MPC) in response to spatiotemporal incentive signaling from the electric grid in the form of locational marginal real-time pricing. Both applications fall under the realm of decision support and could be considered forms of cyberphysical recommendation systems; i.e., occupants interacting with the building as HITLs, and operators interacting with the human machine interface for proactive building health and energy performance.

OpBEMs are thus derived from the practical considerations needed to address the lion's share of energy consumption; i.e., on a daily basis for supervisory control purposes, or cumulative over various retrospective horizons (include over a building's lifetime). Planning for building operational performance, focused on energy consumption or otherwise, results in a specific building's optimum being achieved at the behest of decision-makers. The mathematical models, in which causality may be known or is the subject of learning, can be purely physical, referred to as a white-box model, or purely statistical or data-driven and referred to as a black-box model. The two combined, in some form, are referred to a gray-box approaches and have garnered considerable interest because they balance model complexity and information. An alternative explanation to gray-box models is that they allow uncertainties to be lumped together, but subjectivity still filters the inferences drawn from the model because the modeler must specify how to do so. All modeling efforts may be described as some transition "from dark to light," with inputs mapping outputs through some

(unknown or unobserved) complex functional relationships such that the end-user may be ignorant of the black-box details but seeks a white-box causal explanation.

3.2.2 Tool Abstraction for One-of-a-Kind Buildings

Examining the quality of any project output can be defined in terms of the scope, cost, and time involved. Advanced building energy management is proposed here as an OpBEM. Automated decision support is thus phrased in terms of predictive and diagnostic objectives, such that human operators can optimize their actions, or those actions left to machines, according to quantitative optimization and an operator's subjective preferences. The particular system they seek to optimize is likely unique; e.g., one building, a building portfolio, or a heterogeneous cyberphysical business system. A human machine interface tool capable of dealing with any reasonable abstraction of questions relevant to the built environment is sought; particularly, occupant preferences, opinions, or other such beliefs, can serve as data input to some cyberphysical optimization. The uncountable variations of questions involving the built environment suggest unbounded scope, with many project specific difficulties that drive cost and time commitments to capture operational dynamics. These problems are exasperated by severely limited historic data on what may considered ideal, acceptable, or unacceptable energy performance. Automating the process from BAS data will help guide the expensive and precious time of humans, participate in smart grid concerns, and facilitate human productivity and well-being.

Although many building rating systems exist, and sustainable building projects have proliferated, dynamic recommendation systems for building performance are sparse. Solving the primary problem of responsible building design and mitigating a range of negative environmental impacts are, of course, obtained through superior planning and technological advances. The ongoing service to building performance concerns, however, are neglected because the BAS or BMS, central to building operations in the digital era, are largely disregarded because the cost of energy in the United States is far outweighed by the salary of those humans tasked with its upkeep. However, modern concepts of cloud computing and SaaS can enable building services akin to a luxury car's

diagnostics and human-support systems.

Monitoring expenditures beyond electricity, including but not limited to gas, steam, heat, and water, are typically accomplished through EMCS software, and are popular with operators because they are easy to use and can scale the efforts of property managers, commissioning agents, and other stakeholders. Big data analytics can be thought of as the model-free version of such a framework or as a data supplement for building portfolio analysis. That is, from a top-down perspective, building performance optimization can dispatch resources most effectively and guide actionable initiatives. However, in practice there is confusion surrounding how an EMCS can best be leveraged when the cyberphysical optimization involves a one-of-a-kind building. Many building operators are missing out on opportunities to diminish energy impacts while enhancing building performance.

3.2.3 Building Health and Energy Performance

Many building operators do not distinguish between a BAS and an EMCS, using the terms interchangeably, because they spend the majority of their time dealing with practical issues, problems, and faults, leaving little time and budget for experimentation or work flow automation. Although the BAS automates the control of AHUs, lighting schedules, and dispatch of primary and secondary equipment, it is usually designed to minimum specifications. Sophisticated analytical insight is possible with an EMCS, but constrained by cost (licenses and equipment functionality) and, in practice, the sophistication of users impeding its full potential or power. Continually guiding an operator's decisions is thus nothing completely new; however, an EMCS has typically only been used in major installations like university campuses. Recently, EMCS has been scaled down for smaller installations, primarily driven by sustainability goals associated with building rating systems, and has been used to deliver energy efficiency and enhanced savings. It is argued this entry point is where the aforementioned cyberphysical optimization can commence, but the building health and energy performance is subject to the HMI operators tasked with the problem – and their lack of an appropriate tool.

EMCS systems typically work as an abstraction of the BAS; for instance, corporate headquarters managing the EMCS and satellite locations with control over the BAS. The ability to optimize operations at a building or building portfolio level is typically implemented using a rule-based system. From a top-down perspective, it is possible for the operators to track multiple utility rates, conduct benchmarking energy consumption according to historic data or nearby buildings (while adjusting for weather), and executing building-specific supervisory control to minimize utility demand charges or demand response programs. Identifying opportunities for enhanced building performance, or reducing risks associated with underperforming buildings, is an ongoing task and building health and energy performance naturally interfaces with the BAS/EMCS. Beyond rules, an EMCS that automatically calibrates to the actual building operations, allows the building's performance to be continually assessed without manual or invasive processes. The cost effectiveness of an EMCS can be further enhanced through machine intelligence, physical models, or through manual means, enabling cost reduction or any variety of optimization.

3.3 The Piecewise Development of Tools

The time has come to fully harness building energy models; their extension beyond approaches are termed OpBEMs. With thirty-some years of development, the state of the art in building energy modeling and simulation has matured beyond a tool for design assistance. There has been little exploration of the appropriateness of popular modeling tools for operational concerns, even if a large amount of research has gone into the calibration of energy models according to utility data or similar. This ongoing research interest has long been considered an art as much as a science because of its notorious difficulty. Although guidelines and recommendations exist for building energy model calibration methodology, the lack of uniqueness for a given calibration task does not allow these insights and specifications to be without caveat. It is of great interest to examine model uncertainty over time as data are gathered and insights are sharpened. It is argued that this is inherently a HITL process when dealing with buildings, the people that maintain them, and occupants of varying, subjective opinions.

Decision support for sequential decision making is argued to augment traditional building operations. Building health and energy performance can be systematically assessed on a daily basis to provide decision support — a prerequisite to remote monitoring of building operations such that building services provider groups can be dispatched. It can make recommendations based on 'big data,' extend supervisory control with MPC, or reduce deviations from an ideal and steer the building to more suitable energy performance. Fundamental to this work is that there exists a great need for thought revision for OpBEMs. For superior energy performance, the importance of forecasting, diagnostics, and the integration of buildings into the larger energy system requires a holistic and universal approach with flexibility and adaptability. With building models frequently discarded upon building construction, they could be used for a myriad of applications, such as updating it with data toward evidence-based inference of probable energy performance scenarios. These models could be used for planning and control. The focus here is a data-driven toolchain for the operational performance analysis and optimization of buildings.

It is argued there are uncountable cyberphysical applications involving smart buildings, and the concepts discussed thus far only scratch the surface; however, most important to progress is the *process*. Figure 3.2 shows the three-step process envisioned for continual improvement of building operations and training of the OpBEM. With regard to the figure, there is an asymptotic approach, or centering on, the 'optimum' or 'ideal' operational performance of a building; this can be thought of as automated control over the building, whether advanced automatic control or automation of HITL processes. The first step is classification, which can be thought of as the best estimate a human expert has on a particular matter using up-to-date data and best-in-class physical models. The second step is the forecasting of operational performance, over the decision period of interest, in a sequential fashion that builds on evidence from the previous step. The third step is retrospective and seeks to understand the errors, which are expected in one sense (e.g., weather forecasts), but are primarily used to reduce uncertainty through learning while inferring possible faulty behavior. The three steps combine to provide decision support for buildings. Later chapters explore these concepts in detail, for particular applications, as enumerated in the figure.

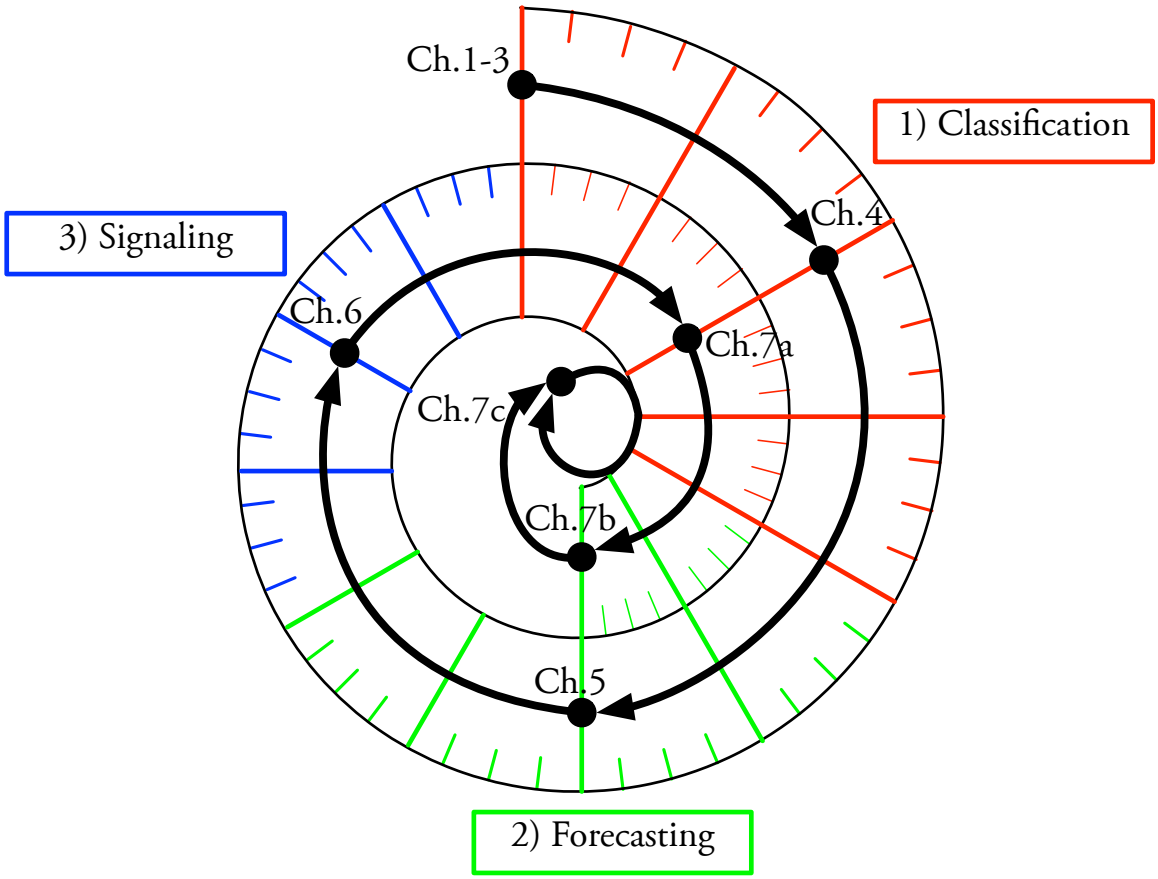

Figure 3.2: The three-step process envisioned as a spiral graph: continual improvement is made in process control by the classification, forecasting, and signaling of operational performance analysis.

Chapter 4

Classification of Commercial Building Electrical Demand Profiles for Energy Storage Applications

4.1 Introduction

Progress has been made in the awareness and implementation of renewable energy systems in a movement toward a sustainable and clean energy future.[1] However, a large opportunity exists for energy efficiency from an integrated, systems engineering perspective. Commercial buildings in particular can benefit from enhanced interactions with the electrical grid; commonly research interest lies in the optimization of indoor environmental quality or single building energy/cost performance. Recognizing that energy efficiency extends beyond building-site boundaries, building operators can work with utilities and/or third-parties to use the building as an asset in various forms of load shifting. Largely neglected in many building-as-an-island performance analyses, such research and demonstration projects are scarce but necessary in growth toward the smart grid.

With commercial buildings being responsible for more than 18% of the annual primary energy consumption in the United States, there are opportunities for various stakeholders to benefit. The appropriate incentive signaling will trigger their active participation in whole-system efficiency and allow utilities to defer capital investments while maintaining operational predictability. Quantifying the shared value among stakeholders is a crucial missing element. Concerns have also been voiced about uncertainties in securing the raw energy resources needed to maintain "business as

[1] Chapter 4 largely derives from Florita et al. [112] and authorship is primarily attributed to Anthony R. Florita; the other authors contributed ideas and edits during their involvement in the Department of Energy's (DOE) Faculty and Student Teams (FaST) Program, where they sought to learn about building energy modeling.

usual," energy conversion emissions leading to harmful environmental impacts, and the intelligent dispatch of renewable energy technologies within the energy infrastructure. Energy storage has been identified as a vital yet largely neglected support for bridging many engineered systems that are orchestrated and optimized for multiple objectives. Proper and cost effective energy storage can be building or utility scale. Confounding the issue is the increased adoption of on-site renewable generation equipment in commercial buildings, such as photovoltaic (PV) panels or wind turbines, because the building acts as both a generator and consumer of electricity. Furthermore, higher levels of renewable penetration with intermittent generation aggravate whole-system instability because supply is not well-matched to demand. Thus, understanding whole-system interactions and mobilizing energy storage for superior dispatch of generation equipment are both essential. Ongoing research seeks a greater functional understanding between commercial buildings and the electrical grid.

4.2 Literature Review

With focus placed on commercial buildings as an asset for whole-system energy efficiency within the larger electrical grid context, the examination of literature followed suit. The natural shared value in this supply-demand system can be expressed in monetary terms. By placing the true value (cost) of unit energy conversion on unit energy pricing the utilities incentivize demand response. With proactive energy management this saves consumers utility costs and generators inefficient expenditures. A number of building-related studies addressed these points in various forms. Investigations of building thermal energy storage have mostly focused on thermal mass control and chilled water or ice storage. Electrochemical storage can also be helpful for renewable technologies and load shifting applications, but due to its relatively high cost current applications are in their infancy.

Thermal mass control harnesses the inherent thermal capacitance of the building for load shifting through supervisory control of zone temperature setpoints and existing heating/cooling equipment. Lee and Braun [113] experimentally evaluated various strategies for a five-hour demand-

limiting period, reductions in peak cooling load up to 30% were reported with minimal disturbance to occupant comfort. A simulation study by Henze et al. [114] investigated the sensitivity of parameters governing the building thermal mass control process and developed a model predictive control environment for optimizing cost savings.

Exploiting the sensible heat of a water body has long been a strategy for implementing thermal storage. A tank system can be integrated into a building's existing hydronic equipment, and finds its greatest use on campuses because of the economy of scale. Bahnfleth and Joyce [115] showed improvements in a university's cooling plant efficiency and achieved demand cost savings with a nearly $17,000\ m^3$ stratified chilled-water storage. Sohn et al. [116] shifted 3 MW of electrical demand from on- to off-peak for an Army installation using stratified chilled-water storage.

With ice storage the energy stored in latent form is a magnitude greater than the amount in an equal volume of chilled water, i.e., per unit volume the "cool" energy stored is about 83.3 MJ/m^3 with ice and about 8.3 MJ/m^3 with liquid water. This can be crucial for commercial building applications where space is usually limited. Furthermore, with storage limitations the control of systems is of upmost importance for the most (cost) effective load shifting. Henze [117] surveyed research in the control of ice thermal energy storage, including conventional, near-optimal, and optimal control. Braun [118] developed a near-optimal control method for charging and discharging an ice storage system in the presence of real-time pricing, the simplified method was shown to be within 2% of the optimal solution.

Energy storage for intermittency issues and the integration of renewable energy has become increasingly important, especially for utility-scale applications. Wind- and solar-generated electricity continually constitutes a larger portion of state renewable portfolio standards. Cavallo [119] addressed these concerns in the discussion of electrochemical, flywheel, compressed air, and hybrid utility energy storage approaches; harnessing commercial buildings as a capacitive element was not considered. Moreover, lacking need for energy conversion, electrochemical energy storage appears attractive to utilities. Kempton and Tomić [120] considered vehicle-to-grid power, which employs electric-drive vehicles (e.g., battery, fuel cell, or hybrid) to interact with specific electric markets.

Strategies and business models were considered, having repercussions for commercial buildings as plug-in electric vehicles become more common.

4.3 Methodology

This research developed and applied unsupervised classification of commercial buildings according to their electrical demand profile. The nonstationary nature of such a time series embeds diurnal swings from occupancy cycles, and contains both higher frequency (e.g., hour-to-hour) and lower frequency (e.g., seasonal) electrical demands associated with weather and climate, respectively. A Department of Energy (DOE) commercial building database [121] was employed. The database contains hourly electrical demand profiles representing the United States commercial building stock as detailed in the 2003 Commercial Buildings Consumption Survey (CBECS) [53] and as modeled in the EnergyPlus building energy simulation tool [122] for a typical meteorological year. Griffith et al. [123] described how the whole-building simulation tool was used for bottom-up modeling of the entire United States commercial building sector. Results from the 4,820 models showed good overall agreement with measured site-energy use intensity.

As fully detailed in following sections, the data were processed in three primary steps. The first step was to perform a discrete wavelet transformation on each electrical demand profile. The second step was to extract unique and descriptive energy and entropy features (absolute and relative) from each wavelet level at definitive time frames. The third step was the application of Bayesian probabilistic hierarchical clustering of the features to classify the buildings in terms of similar patterns of electrical demand. The process yielded a categorized and more manageable set of representative demand profiles, inference of the characteristics influencing supply demand interactions, and a test bed for quantifying the impact of applying energy storage technologies.

4.3.1 Discrete Wavelet Transformation

A common technique to extract time-frequency information from a signal is the windowed Fourier transform, i.e. in transversing a signal, the Fourier transform is combined with a windowing

function of appropriate width to ensure the weakly stationary assumption. The width of the window determines how *time resolution* is traded for *frequency resolution* and vice versa, which is ultimately bounded by the Heisenberg uncertainty principle. Wavelet methods partially overcome this limitation with *multiresolution analysis*. They have proven effective in many signal analysis and engineering applications considering nonstationary transients. Essential discrete wavelet theory and its applicability to the research problem are presented. The reader is referred to Percival and Walden [124] for complete development.

The discrete wavelet transform (DWT) is mathematically stated as:

$$f(x) = \sum_i \sum_j a_{i,j} \psi_{i,j}(x), \tag{4.1}$$

where $i, j \in \mathbb{Z}$, $\psi_{i,j}(x)$ are the wavelet expansion functions, and $a_{i,j}$ are the discrete wavelet transformation (DWT) coefficients of $f(x)$. The coefficients are calculated as:

$$a_{i,j} = \int_{-\infty}^{\infty} f(x) \psi_{i,j}(x)\, dx. \tag{4.2}$$

The wavelet expansion functions are generated through scaling and translation of the mother wavelet:

$$\psi_{i,j}(x) = 2^{-i/2} \psi(2^{-i}x - j), \tag{4.3}$$

where i and j are the scaling and translation parameters, respectively. There are numerous mother wavelet families, e.g. Haar, Daubechies, etc. They are not unique and must satisfy a few mathematical conditions, the most notable being the multiresolution condition. Ensuring the multiresolution condition is satisfied allows the successive transformation of the signal at discrete time-scale ($\approx$ time-frequency) while allowing perfect reconstruction of the original signal. A *scale* is the wavelet's width at level j, an approximate Fourier period or inverse frequency. High scale corresponds to a global view of the data (e.g., the yearly electrical demand profile shape) and the low scale corresponds to detail information (e.g., hourly electrical demand fluctuations from HVAC equipment operation).

Mallat [125] was the first to devise a fast and pyramidal algorithm, deriving the relationship to filter banks. Figure 4.1 shows how the DWT is computed in practice and as accomplished herein.

The $J-$level DWT can be calculated as:

$$\begin{aligned} f_0(x) &= \sum_k a_{0,k}\phi_{0,k}(k) \\ &= \sum_k a_{J+1,k}\phi_{J+1,k}(x) + \sum_{j=0}^{J} d_{j+1,k}\psi_{j+1,k}(x), \end{aligned} \tag{4.4}$$

with $\phi(x)$ being the scaling function, related to the mother wavelet, and where the coefficients $a_{0,k}$, $a_{j+1,n}$, and $d_{j+1,n}$ are either known or calculated from the previous wavelet level according to:

$$\begin{aligned} a_{j+1,n} &= \sum_k a_{j,k}h(k-2n) \\ d_{j+1,n} &= \sum_k a_{j,k}g(k-2n). \end{aligned} \tag{4.5}$$

At level j, the wavelet coefficients a_j and d_j are termed *approximation* and *detail* coefficients, respectively. From the figure, in terms of a filter bank, g and h are half band high-pass and low-pass filters with corresponding filter coefficients. In addition, the figure illustrates that due to the filtering operation (at each level j) half the samples can be discarded according to the Nyquist criterion. *Downsampling* the signal by two (removing every other datum), denoted as $\downarrow 2$, leaves half the number of points and double the scale. Because the filter removes half the frequencies this can be interpreted as losing half the information – the resolution of the signal is halved. In general, the DWT analyzes the signal at different frequency bands, with different resolutions, by transforming the signal into coarse approximation and detail information – the coarse approximation is further filtered as possible.

Each level j of the DWT reduces the time resolution by half since the resultant time series is half the previous level's number of samples. The frequency resolution doubles since the signal's frequency band is half of the previous level's bandwidth. This reduces the uncertainty in the frequency by half. As shown in Figure 4.1 the process is repeated until no more data points remain, with each process iteration sequentially named from level 1 to J. Each level's bandwidth F is labeled to shows the reduction of time resolution and increase in frequency resolution.

When the signal strongly correlates with a wavelet of particular scale and time, large amplitudes are observed. Figure 4.2 shows the first three levels of a DWT with the Daubechies-4 wavelet

utilized in this research; 13 decomposition levels are possible with the 8,760 data points of each electrical demand profile. The bottom-most subplot is an example of a yearly (sampled hourly) electrical power demand profile $x[n]$.

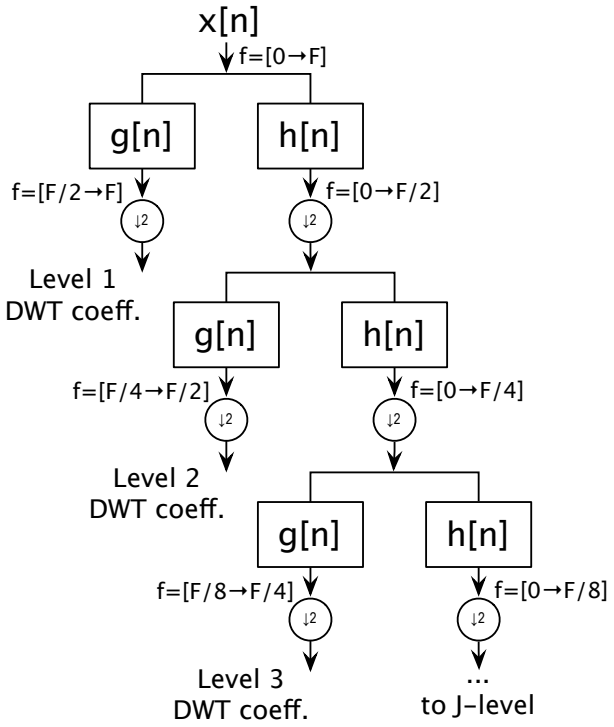

Figure 4.1: The Discrete Wavelet Transformation (DWT) performed using a filter bank for multiresolution analysis of the electrical demand profiles.

As compared to the Fourier transform, the time localization of the frequencies will not be lost with the DWT; however, the time resolution is dependent on the level. Furthermore, more information in the signal resides at higher frequencies and this is where time localization is more precise because of more samples. At low frequencies the frequency resolution is better, but has low time resolution because of the limited number of samples. Frequency bands (i.e., scales) of little importance will have low amplitudes and in many cases can be discarded with a small loss of information. It should be noted that only a portion of the temporal information from the DWT was used herein because of the nature of unsupervised classification (described below); however, the

DWT was selected over Fourier methods so ongoing research could further harness time-frequency information in perturbation studies of individual buildings or building classes – allowing more focused studies of electrical demand at critical time periods for the development of functional relationships.

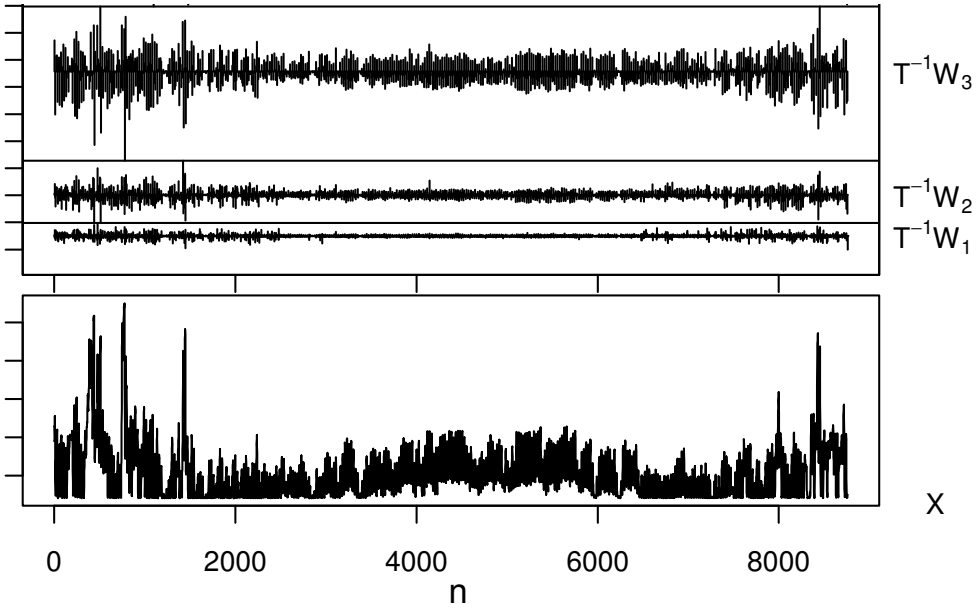

Figure 4.2: The first three levels of a DWT using the Daubechies-4 wavelet: the magnitude at a given time indicates the correlation strength at that level, units are not included for visual clarity.

4.3.2 Feature Extraction

Successfully classifying the commercial building according to their electrical demand profiles rested on the expressivity and the independence of the features extracted from the various wavelet levels. Reducing the dimensionality of features is crucial for any classification algorithm because of the "curse of dimensionality." A number of statistics and/or metrics were considered, but the following signal energy and entropy features were found most informative for inferences and effective within clustering algorithms. Furthermore, due to the broad nature of the supply-demand research concerning (commercial) building-grid interactions the primary focus was: 1) how energy consumption/demand is patterned, and 2) the extent of disorderedness from patterned behavior.

Classification results will allow specification of energy storage technologies according to where the signal energy lies and for smoothing disorderedness during critical time periods of electrical power demand.

4.3.2.1 Relative Wavelet Energy

With the wavelet coefficients determined, termed $c_j[k]$ here for generality, the signal energy at each resolution level $j = 1, 2, \ldots, n$ was:

$$E_j = \sum_k |c_j(k)|^2. \tag{4.6}$$

The total signal energy was calculated as:

$$E_{total} = \|x[k]\|^2 = \sum_j \sum_k |c_j(k)|^2 = \sum_j E_j. \tag{4.7}$$

For each resolution level the relative wavelet energy was calculated as:

$$E_{j,rel} = \frac{E_j}{E_{total}}. \tag{4.8}$$

This ratio indicates the *relative wavelet energy* contained in each time-scale resolution.

As illustrated in Figure 4.3, a histogram $\{E_j\}$ for each commercial building was plotted to compare and contrast the electrical demand profile components for each wavelet level. For classification purposes all possible wavelet levels were considered – 13 separate signal energy features were extracted from each commercial building's electrical demand profile. The signal energy features were informative because of their relative sense, i.e., building size (correlating with demand magnitude) and climate (geographic) influences were effectively considered but standardized. From the figure, Building #1 had more pronounced relative wavelet energy in level 4 and the latter wavelet levels; level 4 roughly corresponds to diurnal variations and levels 12 and 13 with seasonal imbalance. In comparison, Building #4820 contained more relative wavelet energy in the middle wavelet levels (i.e., levels 5 through 9). Observing the original demand profile signals, only seasonal imbalances were obvious.

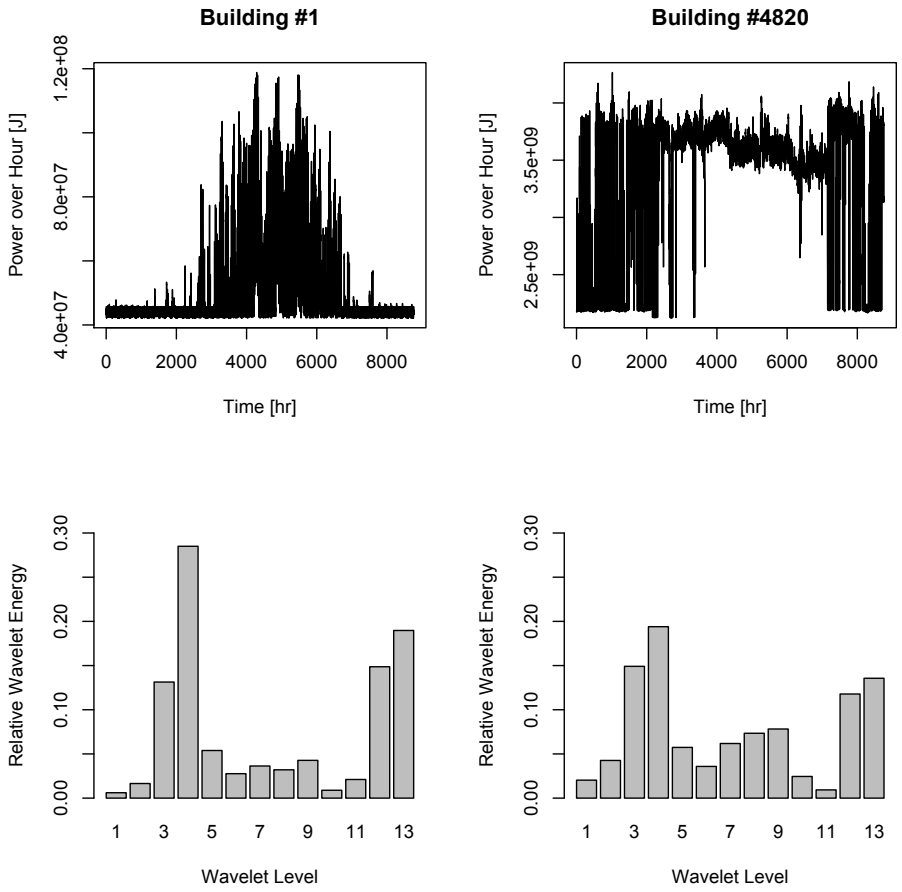

Figure 4.3: Two typical building electrical demand profiles and their relative wavelet energy comparison.

4.3.2.2 Total Wavelet Entropy

Shannon entropy provided a measure of information contained in the signal:

$$H_{wt}(E) = -\sum_{j} E_{j,rel} \cdot ln[E_{j,rel}]. \tag{4.9}$$

In this context it measured the degree of signal disorder, providing information about the underlying dynamical processes. Well-ordered signals are more predictable and can be represented with a fewer number of wavelet levels. Complex signals, usually embedded with stochastic influences, are less predictable and a larger number of wavelet levels are required for representation. Shannon entropy provided a means for quantifying these observations and its calculation was termed the *total wavelet entropy* of a given electrical demand profile.

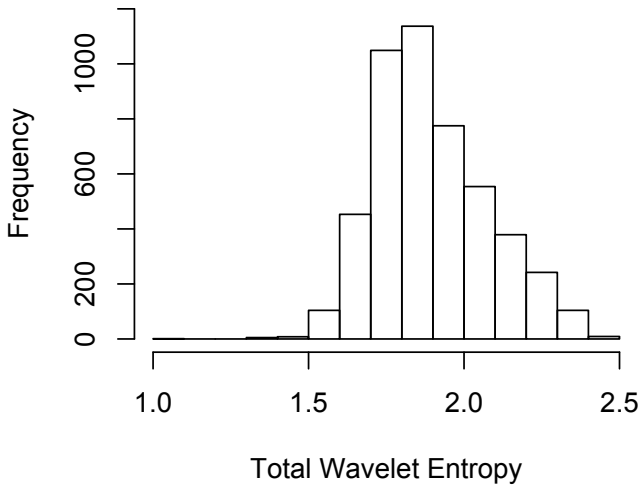

Figure 4.4: Frequency plot of total wavelet entropy for each of the 4,820 commercial buildings.

The frequency plot of Figure 4.4 represents the total wavelet entropy for all 4,820 commercial building electrical demand profiles. That is, $H_{wt}(E)$ is one feature extracted for classification purposes from each commercial building's yearly electrical demand profile. From the database, the minimum, median, and maximum total wavelet entropies of 1.072, 1.865, and 2.433 were obtained for Buildings #1040, #3827, and #1361, respectively. Furthermore, seasonal total wavelet entropies (i.e., winter, spring, summer, and fall – defined by solstices and equinoxes) were extracted as four separate features.

4.3.2.3 Relative Wavelet Entropy

Signal disorder in a relative sense was calculated as:

$$H_{wt}(E_b|E_a) = -\sum_j E_{j,rel,b} \cdot ln\left[\frac{E_{j,rel,b}}{E_{j,rel,a}}\right]. \tag{4.10}$$

Distinctness of commercial building electrical demand behavior was quantified through this *relative wavelet entropy* calculation.

Two different varieties of relative wavelet entropy were considered: the *outer* calculation, which relatively compared independent entropy distributions, and the *inner* calculation, which examined local entropy distributions relative to the signal itself. For the outer version, the relative

wavelet entropy was conditionally calculated against the minimum, median, and maximum total wavelet entropies and used as three separate extracted features. For the inner version, the relative wavelet entropy was conditionally calculated against the time-sequential seasonal total wavelet entropy, i.e., spring|winter, summer|spring, fall|summer, and winter|fall, and used as four separate extracted features.

4.3.2.4 Classification Feature Set

In total 25 features were extracted from each electrical demand profile – those found in testing to most effectively distinguish clusters – and an approximate balance of extracted signal energy and entropy features was noted. This includes 13 relative wavelet energy features, 5 total wavelet entropy features (1 yearly and 4 seasonally), and 7 relative wavelet entropy features (3 moment-based and 4 seasonally). The Bayesian Hierarchical Clustering algorithm summoned the feature set for the purposes of classifying the commercial building electrical demand profiles.

4.3.3 Bayesian Probabilistic Hierarchical Clustering

Hierarchical data clustering is usually a bottom-up agglomerative algorithm and a common method for unsupervised learning or classification. Typically, a distance metric is specified to assess the relative "closeness" of clusters and a threshold measure is used for stopping the clustering process. Each datum (or feature vector) is initially its own cluster and then the two closest clusters are iteratively merged. The process yields a final number of clusters (or classes) of data containing similar features. Although choosing definitive features is difficult, the task of specifying meaningful distance and threshold metrics presents further obstacles. These limitations can be partially overcome with data standardization, bootstrapping (resampling) methods, and clever project-specific metric definitions.

Difficulties with the classical approach have been overcome with recent advances. Heller and Ghahramani [126] employed the powerful Bayesian perspective for probabilistic hierarchical clustering, where marginal likelihoods are used in assessing which clusters to merge and to avoid

over-fitting the problem; this was the chosen approach for the unsupervised classification problem examined in this research. Advantages of the approach are summarized in four main points: 1) a probabilistic model is defined solely through data and used for computing the predictive distribution of (unseen) data for cluster inquiries, 2) a model-based criterion is used to assess whether or not to merge clusters instead of ad-hoc distance measures, 3) Bayesian hypothesis testing is used in decision making: whether to merge clusters or not and the depth of the tree, and 4) the algorithm is fast and has been proven to work with real-world (biological) data.

Using the Bayesian approach, all data (feature vectors) are initially their own cluster (subtree) and two hypothesis are considered in the cluster merging process. For data in a possible cluster, D_c, the first hypothesis, $\mathcal{H}_1^c$, is that all data were generated, independent and identically distributed (IID), from the same probabilistic model, $p(x|\theta)$, where parameter θ is unknown. A prior distribution over the model parameters need to be specified, $p(\theta|\beta)$, with hyperparameter β; a prior represents a priori knowledge of the parameter, but if it is unknown a uniform distribution is used. The probability of D_c under $\mathcal{H}_1^c$ is calculated as:

$$\begin{aligned} p(D_c|\mathcal{H}_1^c) &= \int p(D_c|\theta)p(\theta|\beta)d\theta \\ &= \int \left[\prod_{x^{(i)} \in D_c} p(x^{(i)}|\theta) \right] p(\theta|\beta)d\theta. \end{aligned} \tag{4.11}$$

This model-based criterion evaluates the degree-of-plausibility that some data belongs to a given cluster. The alternative hypothesis, $\mathcal{H}_2^c$, is that the data D_c has two or more clusters in it. The probability of the data under $\mathcal{H}_2^c$ is a product over the subtrees T_i and T_j. Synthesizing the data probabilities under both $\mathcal{H}_1^c$ and $\mathcal{H}_2^c$, it can be shown that the marginal probability of the data in a tree T_k (recursively defined) is:

$$\begin{aligned} p(D_c|T_k) &= p(\mathcal{H}_1^c)p(D_c|\mathcal{H}_1^c) \\ &+ (1 - p(\mathcal{H}_1^c))p(D_i|T_i)p(D_j|T_j). \end{aligned} \tag{4.12}$$

The first term considers the hypothesis that there is a single cluster in D_c, and the second term sums over all other clusterings of data in the tree. Bayes's theorem is applied to determine the

posterior probability of the merged hypothesis:

$$p(\mathcal{H}_1^c|D_c) = \frac{p(\mathcal{H}_1^c)p(D_c|\mathcal{H}_1^c)}{p(\mathcal{H}_1^c)p(D_c|\mathcal{H}_1^c) + [1 - p(\mathcal{H}_1^c)]p(D_i|T_i)p(D_j|T_j)}. \tag{4.13}$$

The quantity $p(\mathcal{H}_1^c|D_c)$ is used to decide which two trees to merge and the final structure of the hierarchy; the tree is cut at points where $p(\mathcal{H}_1^c|D_c) < 0.5$. That is, $p = 0.5$ is the division between "yes" and "no" for the cutting of the tree. The Bayesian hierarchical clustering algorithm is efficient, basing clustering solely on the probability of observed data features.

4.4 Results

This research was programmed in a publicly available computational environment [127] using publicly available data [121] for the unsupervised classification of commercial buildings based solely on their electrical demand profile. Using the methodology outlined, the initial dataset of 4,820 buildings was reduced into 114 classes for a compression of greater than 97%. The process yielded a categorized and more manageable set of representative electrical demand profiles, inference of the characteristics influencing supply-demand interactions, and a test bed for quantifying the impact of applying energy storage technologies.

Due to the large number of buildings (4,820) the results of the clustering were nearly impossible to represent in the classic dendrogram format. Nonetheless Figure 4.5 presents the complete clustering process, while neglecting inclusion of the building labels due to significant overlap, so the reader might visualize the 4,820 buildings merging into the final 114 classes according to the extracted features. Figure 4.6 displays a limited portion of the full dendrogram, classes 8 and 9, demonstrating the merging hierarchy of the buildings within the class. The red dashed lines on the dendrogram represent merges that the clustering algorithm does not make, while the number displayed represents the log odds for merging.

Manual inspection of each class represented a significant challenge. Each building has tens of thousands of variables and parameters, and significant results may be overlooked. Because of

this challenge, the approach taken herein utilized graphical techniques for the sake of presentation, as well as simplifications appropriate to the CBECS data [53] utilized. That is, only high-level characteristics within each class were considered for condensed results presentation and for the demonstration of general results. For illustration purposes, three primary characteristics were focused on to demonstrate a subset of possible inferences, as well as to show relevant characteristics influencing the classification: the climate, the principal building activity, and the building size. Finally, a simple example is given to exemplify one of many possible ways the identified building classes might be utilized with respect to energy storage analysis.

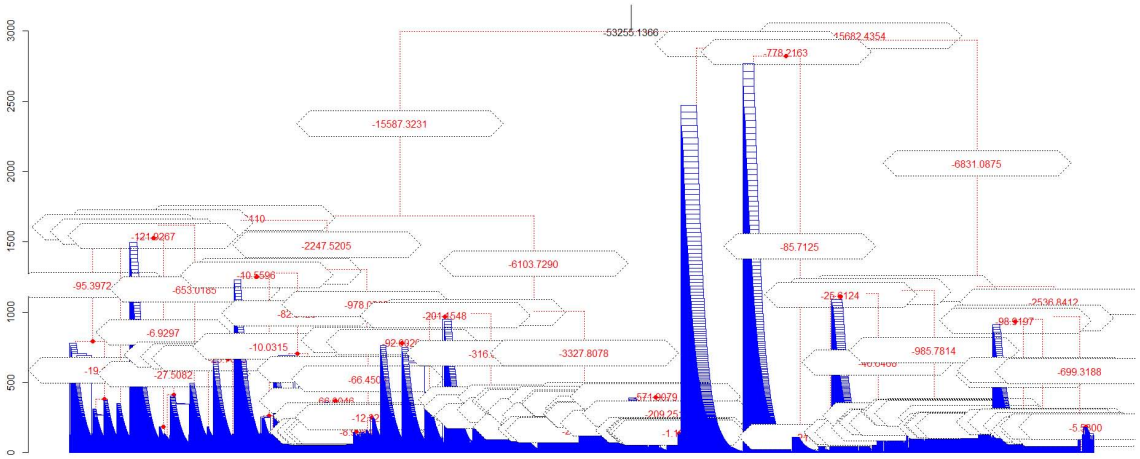

Figure 4.5: Dendrogram of the clustering process for the united states commercial building stock: visualization of the 4,820 buildings merging into the final 114 clusters (building classes) according to the 25 features extracted from the electrical demand profiles; for visual clarity building labels not included.

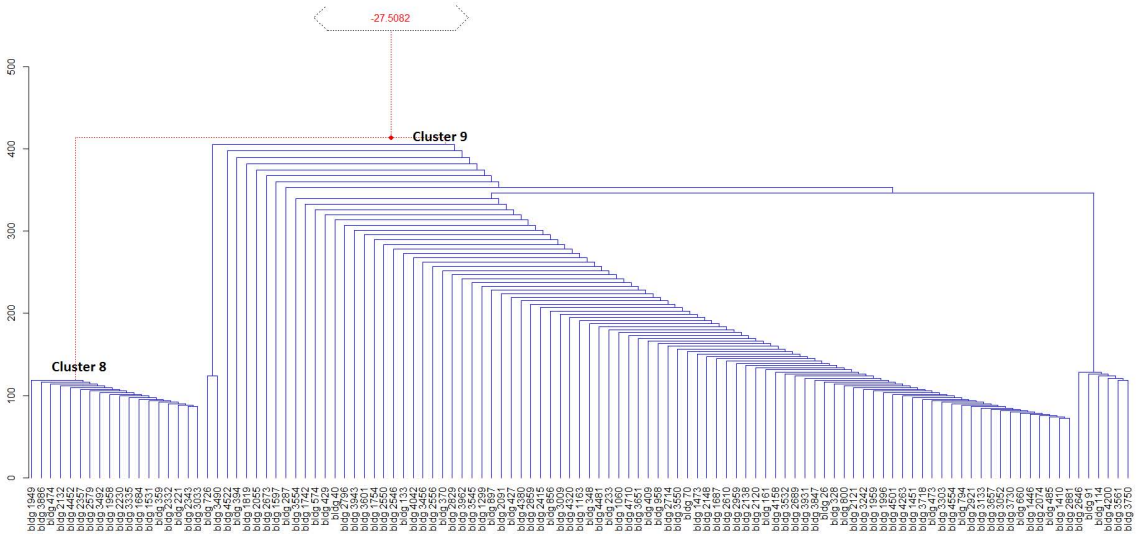

Figure 4.6: Limited portion of dendrogram (with focus on clusters 8 and 9) demonstrating merging hierarchy; red dashed lines are merges not made, while the number displayed indicated log odds for merging.

4.4.1 The Influence of Climate

The first characteristic that was investigated for its influence on the classification of buildings was the climate region. The CBECS database utilizes five climate regions based on the number of cooling and heating degree days (CDD and HDD respectively). Figure 4.7 shows the heatmap based on the climate zone used in the modeling and the resulting classes. The darkest red rectangles represent classes with a very small number or zero buildings for the given climate region being represented in the class, while white (or bright yellow) represents a class with a large number of buildings from a given climate zone. For the sake of discussion, focus was placed on a few classes to demonstrate the results of the technique. For example, classes 48 and 90 are highly represented by buildings from climate zones 4 and 1, respectively. Whereas, classes 72 and 103 are most represented by buildings in zones 3 and 1, respectively, but also were significantly represented in other climates.

From the heatmap it can be seen that climate zone does not necessarily represent a strong driver for classification. This implies that energy storage control strategies might apply irrespective of climate and are modified solely by magnitude to ensure coverage of signal energy and entropy

features. The caveat to this line of thinking is that the utility rate structure for a given building will vary across the country and will drive the energy storage control strategy. However, it is feasible that multiple buildings serviced at a given bus might be analyzed for energy storage opportunities and the appropriate demand profile from each respective class could simply be added. Overall the heatmap showed climate regions associated with specific classes did not demonstrate any conclusive patterns. A variety of reasons within the modeled database could lead to this: first, multiple weather files were used to model buildings within a given climate zone [123]. This leads to more variety in weather within a given climate zone. Second, differing building sizes and types could lead to significantly different operating patterns within a given climate zone.

4.4.2 The Influence of Building Type

The second characteristic investigated was based on the principal building activity. The CBECS database classifies buildings based on their major usage with such categories as office space, retail, and inpatient health care to name a few. It can be seen in Figure 4.8 that building type often leads to selection of a load profile for a given class. For example classes 71 and 104 are predominantly filled with buildings used for inpatient health care. It is of interest to determine: if principal building activity is so dominant why these two clusters are not merged into the same cluster? A reason which will be addressed in the next subsection. Additionally classes 48 and 91 were not dominated by one building type but share a variety of building types, but these classes were made up of buildings primarily from one climate zone. On the other hand, offices, which are also the largest set of buildings by building type within the database, were dominant in a variety of classes. Interestingly the majority concentrations were located in adjacent classes, demonstrating their likely appearance in higher level merges. Although the Bayesian hierarchical clustering routing does not consider these merges to be highly probable, it did indicate the potential for similar behavior.

4.4.3 The Influence of Building Size

Square footage was also a highly revealing result, particularly when the discussion includes the effects listed in the previous section. For example it was noted that classes 71 and 104 were primarily made of buildings used for inpatient health care but were not merged. When this result was combined with Figure 4.9 it could be seen that these two classes were not merged because of the difference in square footage. Class 71 was primarily made of smaller health care buildings with class 104 being made up of the largest two square footage categories in the data set; the difference in classification is likely attributed to vastly different HVAC equipment, even though the (standardized) occupancy and usage patterns of the buildings are similar. Turning the attention to classes 48 and 91 again, it can be seen that these classes were made up of mixed results based on square footage.

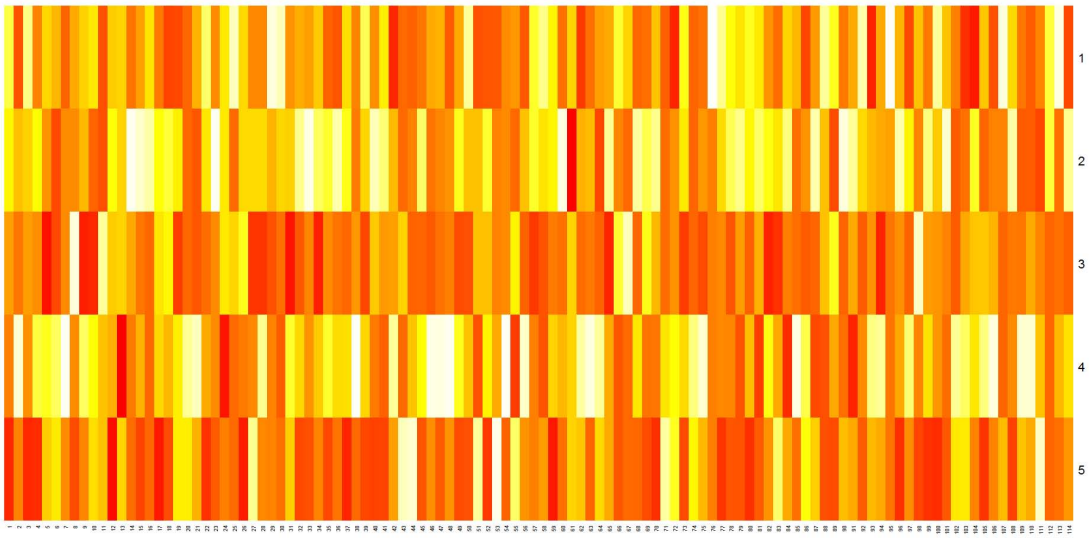

Figure 4.7: Clustering heatmap based on united states climate zones: CBECS uses five climate regions based on cooling and heating degree days. Darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number.

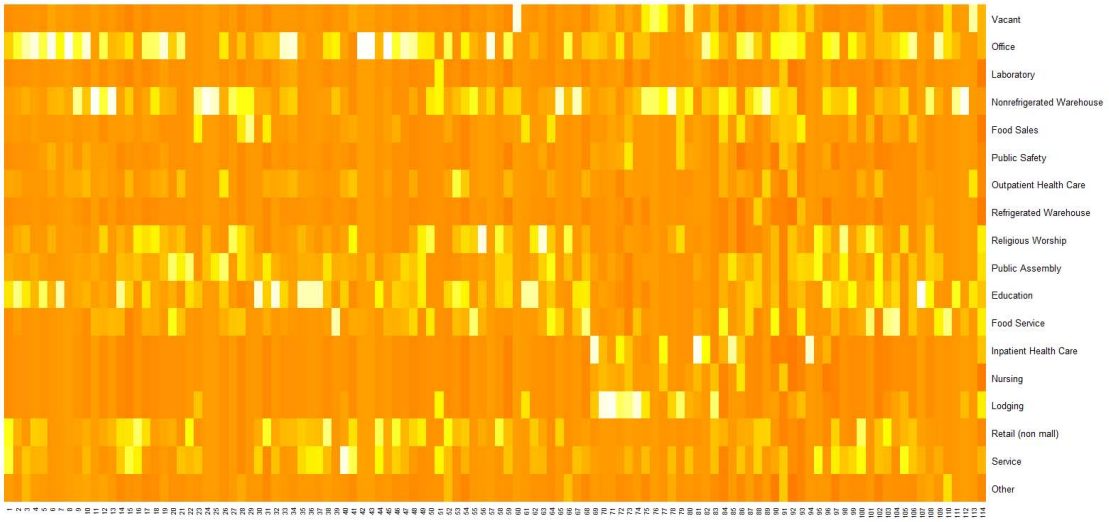

Figure 4.8: Clustering heatmap based on principal building activity: CBECS classifies buildings according to their major usage, e.g. office, retail, etc. darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number.

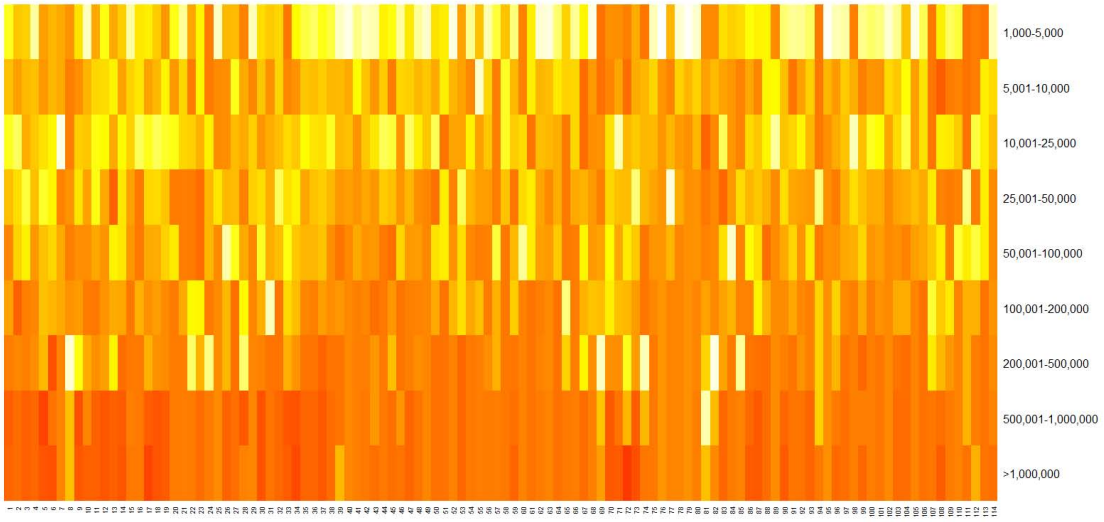

Figure 4.9: Clustering heatmap based on square footage. Darkest rectangles represent a cluster with a small number of buildings (or zero), while white rectangles represent a cluster with a large number.

4.4.4 Perceived Use of the Commercial Building Classes

It should be noted the unsupervised classification represents only the first-level, top-down approach to uncovering functional relationship between a building electrical demand and its governing parameters and/or variables. The identified classes of commercial building according to their demand profile have two perceived uses with respect to demand response and/or energy storage:

Unit Commitment and Dispatch Models —are used to determine scheduling for start up and/or shut down of power plants in a system area of interest, i.e. usually across multiple states. Mixed integer programming is used to solve the commitment problem with forecast weather and demand, with each unit having unique characteristics and constraints (e.g. up and down times, ramp rates, scheduled maintenance, etc.), and the solution is modified and dispatched on a shorter horizon according to updated and hopefully more accurate information. Since even percentage improvement in demand forecasting can save tens of millions of dollars on a yearly basis, there is a great need to better understand electrical demand at a particular distribution bus and better predict it. The aggregation of the appropriate building demand profile classes, where perhaps transmission is constrained and thus power is costly, can provide a starting point for further analysis. This can help in decision making with respect to demand response incentives or the addition of energy storage.

On a Building Level —the understanding of how building parameters and/or variables influence the demand profile is incomplete. The unsupervised approach was helpful for uncovering correlations and natural grouping. It is noted that the bottom-up, deterministic modeling and simulation approach used to produce the commercial building demand profiles[123], albeit the state-of-the-art [122], provided only an estimation of nominal behavior and neglects considerable uncertainty. However, it is the authors' belief that since the database[121] was derived from an elaborate statistical analysis of the building stock [53], it provided a good starting point for this research. It is apparent there are uncountable combinations of data analyses and regressions that could be performed to better understand functional relationships – even without considering the

uncertainty of each parameter and/or variable or the accuracy of the simulation models themselves. Furthermore, it is apparent building-level thermal perturbations (e.g. thermal energy storage and release) will influence the demand profile and any given building might shift classes.

4.4.5 A Classified Building Energy Storage Example

An example is given on how a building class might be used to aid the energy storage design for a particular building. The first step would be to examine all buildings within that class to better understand possible trends. When examining a single commercial building for the application of energy storage, one might be interested in a specific energy storage technology, have a good understanding of the approximate storage capacity needed, and optimize the design according to a particular utility rate structure to save operating costs.

The representative commercial building demand profile from class #1 (scaled by the magnitude of peak demand from the first building in the class) is considered for the month of July. In addition, on-site PV panels were added such that the solar fraction was equal to unity if all energy can be stored, i.e. the total energy needed to operate the building was equal to the PV production but the supply and demand do not necessarily match. Figure 4.10 shows the building demand profile, the PV production, and the hourly difference, respectively. The necessity for storage is evident during periods when PV production is high but the building demand is low; this is especially notable on the weekends. This is a problem of a "highly energy efficient building" where simply adding an abundant amount of PV without regard to the whole system efficiency leads to problems. That is, it is possible the power peaks flowing back into the grid make the whole-system efficiency worse. Furthermore, the utility is only going to buy power back at a fraction of what it charges for the same quantity.

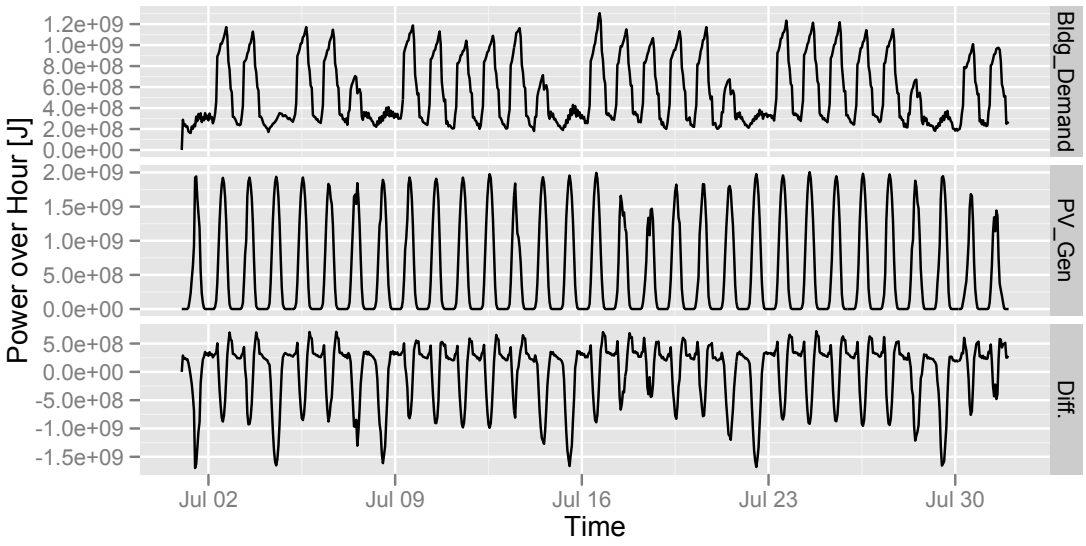

Figure 4.10: The building demand profile, on-site PV generation, and difference for the month of July.

To capture the energy spillage back onto the grid, a generic energy storage was considered. Simple on/off control to (or from) the storage was based on the hourly difference between building demand and PV production. The parameters of interest were: 1) the energy storage size as a fraction of the worst-case design storage capacity, and 2) the maximum charging/discharging rate per hour as a fraction of the design storage capacity. The design storage capacity [J] was specified as the greatest diurnal difference between building demand and PV production during the month of July, i.e. the theoretical ability to store all excess production on the worst-case July day. The maximum charging/discharging rate parameter allowed the design storage capacity to go from empty to full in one hour or vice versa. Charging and discharging efficiencies were assumed to be 90%. Temperature degradations were not considered. Although the parameters were for a generic storage, they represent a range of electrochemical storage technologies.

From Figure 4.11 it is shown that the net energy spillage to the grid was more sensitive to the rate of charging/discharging than the storage size. In both cases, at fractions above approximately 0.50, the solution does not change. With an adequate storage size ($\sim$ 0.35 design capacity) and

charging/discharging rate (~ 0.2 max rates) only 35% of the PV production was sent to the grid and not utilized on site. The maximum (monthly) demand from the grid is also a big concern for building operators because of the expensive charges associated with its peak magnitude; it was interesting to see how much the storage influences peak demand. From the figure, the storage allowed demand savings of 12.5% without dedicated control. In all energy storage studies utility rate structures will have a profound influence, and those results offered here are solely to illustrate classification results.

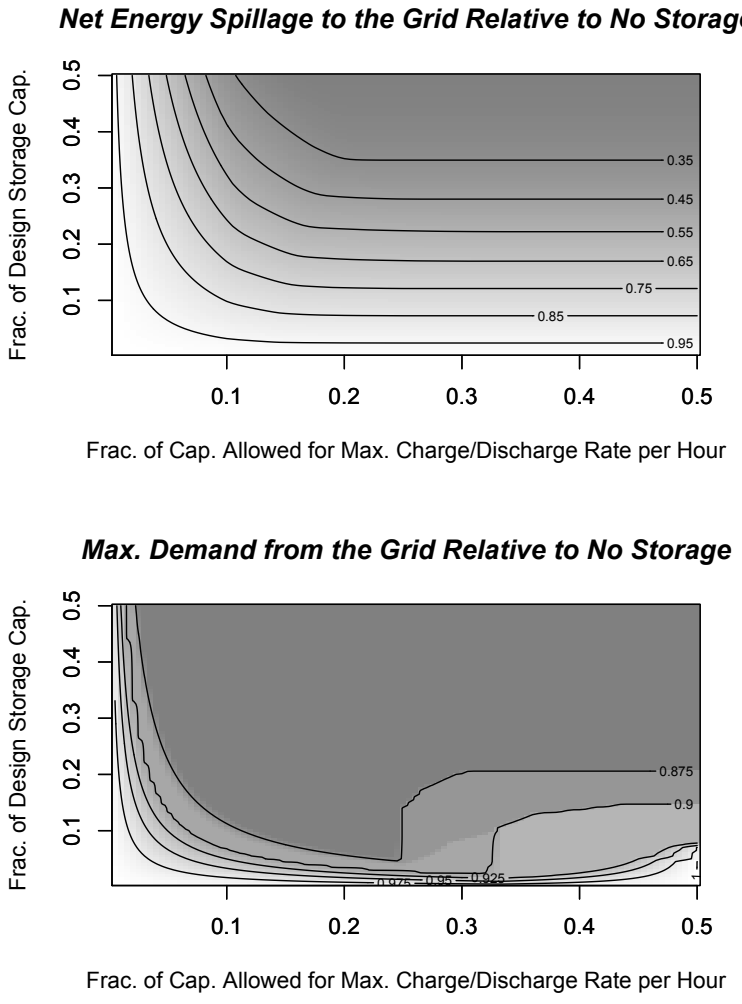

Figure 4.11: Energy spillage and peak demand savings as a function of energy storage parameters.

4.5 Conclusions

Overall the heatmaps presented in Figures 4.7 through 4.9 demonstrated the validity of the technique to classify similar buildings that one might expect to have indistinguishable electrical demands, as well as shed light on unexpected classes. The two-way and potentially higher interactions these figures reveal showed the challenging nature of such a technique but also demonstrated how dynamic building usage can be. This technique and the results presented here represented just the first level of analysis and demonstrated the need for further research within the identified building classes. Particularly, the investigation will be extended to look at the impact of weather from a more detailed level, and the impact of the air-conditioning equipment used. Additionally, investigating if significant variations in classification occur when real weather data are used, versus typical meteorological year data, is of concern. Furthermore, observing what happens to the load profiles when a large suite of advanced technologies, including various on-site generators and energy storage technologies, are included is of great interest. These impacts in particular are highly relevant to the intelligent integration of commercial buildings with renewable generation and energy storage capabilities.

Chapter 5

Comparison of Short-Term Weather Forecasting Models for Model Predictive Control

5.1 Introduction

Model predictive control (MPC) of buildings is highly appealing as it holds the promise of energy and cost savings through intelligent building operation.[1] Progress toward the goal of real-time optimal control of buildings is made through a systematic inquiry of each process component. Although a wealth of research has been conducted in building physics, there remains a great opportunity for the proper integration and control of each of these elements from a supervisory control perspective. Among the obstacles is the accuracy by which the stochastic local weather processes can be predicted. It is necessary to investigate suitable forecasting models to uncover the complexity required for short-term prediction of local weather, while observing the applicability within a MPC framework. Progress is made toward this goal through a systematic evaluation of various climates, weather variables, and forecasting models.

Time series prediction of local weather is crucial for many aspects of energy conservation, economic operation, and improved thermal comfort in commercial buildings. In particular, significant motivation exists for HVAC supervisory control applications, e.g., building thermal mass control to respond to utility pricing signals, increased free cooling through superior economizer control, and thermal load prediction for uniform chiller loading and improved part-load performance. The desire is to optimize operation based on a short-term prediction of weather and building utiliza-

[1] Chapter 5 largely derives from Florita et al. [128] and authorship is primarily attributed to Anthony R. Florita; the other author contributed code and editing, specifically concerning the neural network models.

tion to yield energy and cost savings through the minimization of an objective function over the prediction horizon. Identifying a MPC strategy and determining the forecasting model complexity required will support the causes of the example and additional applications. Taking a generalized data-driven perspective over physical modeling, the relevance ranges from utilization within embedded controllers to high-level setpoint adjustment within the building automation systems (BAS); the availability of accurate dynamic building energy simulation programs further promotes future application of MPC in commercial buildings.

It is well known that government and private institutions alike provide forecasting services involving complex meteorological models and supercomputer computation. The arguments against relying on these services are the five main concerns: availability of all variables of interest, service interruption, service availability, system integration, and the data-driven perspective. First, the presented models can be applied to predicting local solar data such as global horizontal and direct normal insolation, which is not commonly offered by commercial service providers. Second, service interruption could seriously negate the benefits of model predictive control, e.g. with electrical demand charges typically being levied over monthly bill periods, and ratchet clauses extending this charge to future billing periods, any failure of demand limiting strategy caused by a forecasting malfunction could have large cost implications. Third, the availability of hourly (or sub-hourly) weather forecasts is not pervasive. That is, forecast updates required in real-time, necessary for error minimization, are not available through most forecasting services. Error can additionally be introduced through localized deviations, e.g. the grid point for which the forecast was generated does not accurately represent local climatic observations. Fourth, system integration concerns stem from the authors' believe that the "smart building" benefits from an on-site weather station for appropriate model predictive supervisory control. Fifth, a practical, data-driven (on-site) perspective is favored over devising complex (physical) probability models of, say, solar irradiation and/or cloud cover. Therefore, a range of weather forecasting models for MPC have been evaluated: from simple models that only capture the shape of the weather processes to complex models that attempt to capture the stochastic processes and underlying dynamics.

5.1.1 Nomenclature

C	autocovariance function
d	number of days in the past horizon or dimension
df	degrees-of-freedom
dL	execution (prediction) horizon
Fr	Friedman statistic
I_{gh}	global horizontal radiation, Wh/m^2
i, j	indices
k^*	update time instant
L	planning horizon or characteristic profile
MA	moving average
m, n	number of blocks, treatments
Q	studentized range statistic
R	ranking
R_{sd}	relative standard deviation modification
R_{sw}	swing ratio scaling modification
T_{db}	dry-bulb temperature, °C
T_{wb}	wet-bulb temperature, °C
t	specific hour of the day under investigation
x	general variable or weather variable of interest
α	discount factor or level of significance
δ	deviation
μ	mean value
σ	standard deviation
τ	time lag
ϕ	relative humidity, %

χ^2	chi-square statistic
∇	backshift operator
$\{x\}$, $\vec{x}$	vector notation
$\overline{x}$	mean value
$\hat{x}_i$	predicted value

Subscripts, Superscript, etc.

aew	absolute deviation and exponentially weighted
asp	absolute deviation and simple prior
d	number of days in the past horizon or dimension
e	external
ew	exponentially weighted
i, j	indices
k^*	update time instant
M, N	total number of terms
m, n	indices
rew	relative (standard) deviation and exponentially weighted
rsp	relative (standard) deviation and simple prior
sp	simple prior
sw	swing or range
t	specific hour of the day under investigation
∇	backshift operator

5.2 Literature Review

A spectrum of relevant case studies with weather prediction aspects was reviewed. Many studies come from the very active research domain of electrical power forecasting used to optimally dispatch central power generation equipment, with weather forecasting typically influencing the unit commitments. Despite overlaps in forecasting strategies and algorithms, the present concerns limit

the literature review to studies involving commercial buildings with weather forecasting and/or prediction features to corroborate the unique investigation presented herein. In both domains a significant fraction of researchers are convinced that nonlinear forecasting models based on neural networks (NN) provide superior forecasting performance, nonetheless, many have found success using traditional time series analysis. For a detailed presentation of time series analysis and neural networks the reader is referred to Box et al. [129] and Bishop [130], respectively.

Time Series Analysis Models: In forecasting hourly cooling loads, MacArthur et al. [131] used an autoregressive moving average (ARMA) model. A recursive least squares (RLS) algorithm was used for on-line model parameter adjustment with exponential forgetting factor to ensure a maximum covariance between the predicted and actual load. The quality of the load forecast was innately tied to ambient weather prediction because the method used historical data and external forecasts of maximum and minimum values for rescaling the profile.

Seem and Braun [132] developed a model for electrical load forecasting by combining an autoregressive (AR) model for the stochastic part of the prediction task and an adaptive look-up table for the deterministic part. Using an exponentially weighted moving average (EWMA), hourly electrical power demand was determined using previous forecasts and measured values. The deterministic look-up table is often referred to as a cerebellar model articulation controller (CMAC). Its performance showed improvement by introducing electrical load profile temperature dependence, with daily peak values modified according to the maximal daily temperature forecast as provided by the National Weather Service (NWS).

Chen and Athienitis [133] discuss an optimal predictive control methodology for building heating systems in real-time dynamic operation. The system consisted of three loops: a conventional feedback control loop with predictive regulator, a parameter estimator using an RLS technique, and a setpoint optimizer. The weather prediction made use of both historical records and local weather forecasts. Simple rules were devised to modify historic shape factors based on external weather service forecasts.

Henze et al. [134] investigated the impact of forecasting accuracy on predictive optimal control

of active and passive building thermal storage inventory. The short-term weather forecasting models including bin, unbiased random walk, and seasonal integrated ARMA predictors. It was shown that model complexity does not imply accuracy, and progress is built on these results in the context of the MPC framework.

Neural Network Models: Ferrano and Wong [135] used a feed-forward NN software package to predict the next day's cumulative cooling load by mapping hourly (24 input units) ambient dry-bulb temperatures, with results subsequently used in a real-time expert system. In order to allow for adequate generalization and avoid memorization, it was determined that training sets should contain a difference of at least 3% in the temperature patterns. This effectively reduced training time and reduced the prediction error from a maximum of 12% to only 4% for the validation sets.

Kreider and Wang [136] used a feed-forward neural network with nine input units in the prediction of hourly electrical consumption integrated from 15-minute power values. The network was capable of responding to unusual weather phenomena, e.g. during a non-cooling month (December) the weather was unusually hot for two days and the cooling plant had to come on during that time period. The network picked up this unusual condition and predicted the required power consumption for this period. Conventional methods of regression missed this event. It was shown that the quality of the forecast improves when the network is trained on data of the same season for which the prediction is made, i.e. predicting energy consumption for a non-cooling month such as December should be done with a network that is trained on November data rather than data from a typical cooling month such as July.

Gibson and Kraft [137] used a recurrent NN for electric load prediction in conjunction with a thermal energy storage system. The network was trained using electric and cooling load data of an office building monitored over a cooling season. The ability to generalize was noted by the network's ability to achieve similar prediction errors for atypical days as compared to normal days when supplied with detailed occupancy load information. Using a recurrent network architecture allowed the network to not only pick up the steady-state physical processes, but also the temporal

relationships between system states, i.e. dynamic aspects.

Dodier and Henze [138] used neural networks as a general nonlinear regression tool in predicting building energy use data as part of the Energy Prediction Shootout II contest (described below). By combining a NN with a version of Wald's test, the relevance of free parameter inputs were determined. Time-lagged input variables were found by inspecting the autocovariance function. The strategy was the most accurate predictor in the competition.

Previous Comparative Studies in Prediction: The "1993 Great Energy Predictor Shootout", detailed in Kreider and Haberl [139], was a competition in the prediction of hourly building energy data available to world-wide data analysts and competitors alike. The results showed that connectionist approaches were used in some form by all winners, i.e. NN approaches using different architectures and learning algorithms proved to achieve superior accuracy when compared with traditional statistical methods. Kawashima et al. [140] compared hourly thermal load prediction for the coming day by means of linear regression, ARIMA, EWMA, and NN models. The NN models were confirmed to have excellent prediction accuracy and considered by the authors to be superior methods for utilization in HVAC system control, thermal storage optimal control, and load/demand management. The "Great Energy Predictor Shootout II" extended the understanding of prediction methods for building applications, evaluating the prediction of hourly whole-building energy baselines after energy conservation retrofits. The effectiveness of prediction models was compared for the top entries in Haberl and Thamilseran [141]. NN models were shown to be the most accurate in nearly all cases, but unique statistical approaches were shown to compete, in terms of accuracy, with the NN models.

5.3 Description of Analysis

5.3.1 Overview of Short-Term Weather Prediction

To ensure the forecasting models were evaluated for a wide spectrum of geographic locations, eleven climates in the United States, Europe, and Asia were investigated as shown in Table 5.1. For

each city, International Weather for Energy Calculations (IWEC) data was used to train and test the models' forecasting of weather variables. Due to their prevalence as required inputs to dynamic building and component simulation models, and for practical engineering purposes, the following four weather variables have been predicted: global horizontal radiation I_{gh} in units of [Wh/m2], dry-bulb temperature T_{db} [°C], wet-bulb temperature T_{wb} [°C], and relative humidity ϕ [%]. Each forecasting model was used separately and seasonally to predict each of these weather variables for each city.

The seasonal testing months were specified as: March (spring), June (summer), September (fall), and December (winter). To account for realistic, continuous forecasting model operation the previous two months of the yearly IWEC data were used as training data, and the third month as unseen testing data in the appraisal of seasonal forecasting performance: the spring evaluation used January and February for training data and March for testing forecasting performance, the summer evaluation used April and May for training and June for testing, and so forth in the fall and winter seasons. To achieve the research goals through a robust evaluation, the reasoning in the selection of the training and testing strategy follows a tripartite structure: 1.) forecasting models are globally and seasonally evaluated to ensure their performance is not biased by a single climate and/or season, 2.) the specified testing months coincides with weather phenomena typical of the respective season, and 3.) the forecasting models must perform in a data-driven, continuous manner by training on recent (historic) weather data and operating in contiguous months – on simple systems such as embedded controllers.

Table 5.1: Geographic Locations Considered for Short-Term Weather Prediction.

United States	Europe	Asia
Atlanta, Georgia	London, England	Beijing, China
Los Angeles, California	Paris, France	Tokyo, Japan
New York, New York	Stockholm, Sweden	
Omaha, Nebraska	Stuttgart, Germany	
Phoenix, Arizona		

The discrete prediction occurred in hourly intervals, a typical simulation time-step length. The metrics used to assess forecasting accuracy are the coefficient-of-variation (CV) and mean bias error (MBE), both in [%].

$$CV = \frac{\sqrt{1/N \sum_i (x_i - \hat{x}_i)^2}}{1/N \sum_i x_i} \tag{5.1}$$

$$MBE = \frac{1/N \sum_i (x_i - \hat{x}_i)}{1/N \sum_i x_i} \tag{5.2}$$

Where $\hat{x}_i$ is the predicted value of the monitored dependent variable x_i, and the denominator is the mean of the dependent variable over the testing set. In all cases, the metrics are calculated for an execution horizon of $dL = 6$ hours, successively concatenated to encompass the complete testing period. The execution of the prediction constitutes only a portion of the planning horizon of L hours because most of the MA models allow for corrections as time progresses, and the NN models must be compared on the same prediction-task basis.

The length of planning horizon L in a given prediction is primarily a function of the longest time constant of the dynamics to be controlled; i.e., when performing MPC of building thermal mass a longer planning horizon is required over which a strategy can be formulated as compared to the problem associated with faster dynamics such as chilled water plant control involving smaller time constants. Based on Henze et al. [142], planning horizons in excess of 18 hours (better with 24 hours) were suggested for building thermal mass control. On the other hand, the quality of the

planned strategy depends on the quality of the forecasts it rests on. This in turn forces the choice of a planning horizon that accounts for the fact that forecast uncertainty grows with length of time predicted. Guidance is offered for any MPC within commercial buildings, and thus the choice of execution horizon dL would be a function of the specific application. A value of $dL = 6$ hours was chosen with the thermal mass application in mind and recognizing that forecasts tend to be of sufficiently high quality for the next 6 hours. However, both shorter and longer execution horizons could have been chosen for the analysis. It is the belief through considerable testing of reasonable dL time lengths, however, that the findings would not have changed significantly.

5.3.2 Investigated Forecasting Algorithms

A central question was whether the more computationally intensive requirements of neural networks are warranted when compared to the performance of simple and easy to implement time series analysis applied to the cyclical two-stage MPC process of prediction and execution. With various levels of sophistication, both methods were applied in a comparative analysis. Altogether 14 different short-term forecasting models have been proposed, and Table 5.2 provides the naming convention. The respective modeling details are outlined in following sections.

Table 5.2: Labeling of the Various Models Used in the Analysis.

Model	SPMA	EWMA	Neural Network	Ext. Mod.	Abs. Dev. Mod.	Rel. Dev. Mod.
1	x					
2	x			x		
3	x				x	
4	x					x
5	x			x	x	
6	x			x		x
7		x				
8		x		x		
9		x			x	
10		x				x
11		x		x	x	
12		x		x		x
13			FTDNN			
14			NARX			

5.3.2.1 Simple Prior Moving Average (SPMA)

Historic observations are often drawn upon for developing characteristic profiles (forecasts) of weather variables, generically known as past-horizon predictors. The finite past horizon can be variable and may encompass the last day, week, month, or season to generate the forecasts. Equation 5.3 generates a characteristic profile for the next L hour planning horizon on the basis of a simple prior moving average (SPMA) over the past horizon by performing uniformly-weighted averages for each hour of the day:

$$MA_{sp,t}(d) = \frac{1}{d}\sum_{i=1}^{d} x_{t-\nabla i} \tag{5.3}$$

The number of days in the past horizon is indicated by d, the weather variable x is of interest,

t is the specific hour of the day under investigation, and the $\nabla = 24$ hour backshift operator is modified by the index i within the summation to account for the diurnal nature of the weather variables. The investigated SPMA models used $d = 60$; weighting concerns and dynamic adaptation are considered below.

5.3.2.2 Planning and Executing Predictions

Due to the stochastic nature of weather, a measured weather variable datum for a particular hour will deviate from its forecast value. To allow dynamic adaptation to such an event, as the predictor moves through time it accounts for any discrepancy at update time $k*$. It is assumed the deviation will persist for a number of hours into the future. Therefore, the shape of the forecast weather variable is determined by the characteristic profile for a planning horizon L, but the forecast is 'anchored' (described below) to the measured current value. In essence, the predictor assumes persistence of the deviation from the static forecast and adapts to measured values in real-time.

With the dynamic strategy, forecasting of weather variables is accomplished using historic events in the planning horizon L, adaptation according to current measurements at the update instance $k*$, and then executed over dL. At the end of the execution horizon, it is likely that another deviation will exist between the measured and forecast weather variables as shown in Figure 5.1, and the process moves through time. The cyclical two-stage process of policy planning and execution is highly applicable to the requirements of MPC because recent historic data and current weather knowledge are considered, but only short-term forecasts are executed, e.g. an optimal temperature setpoint in MPC.

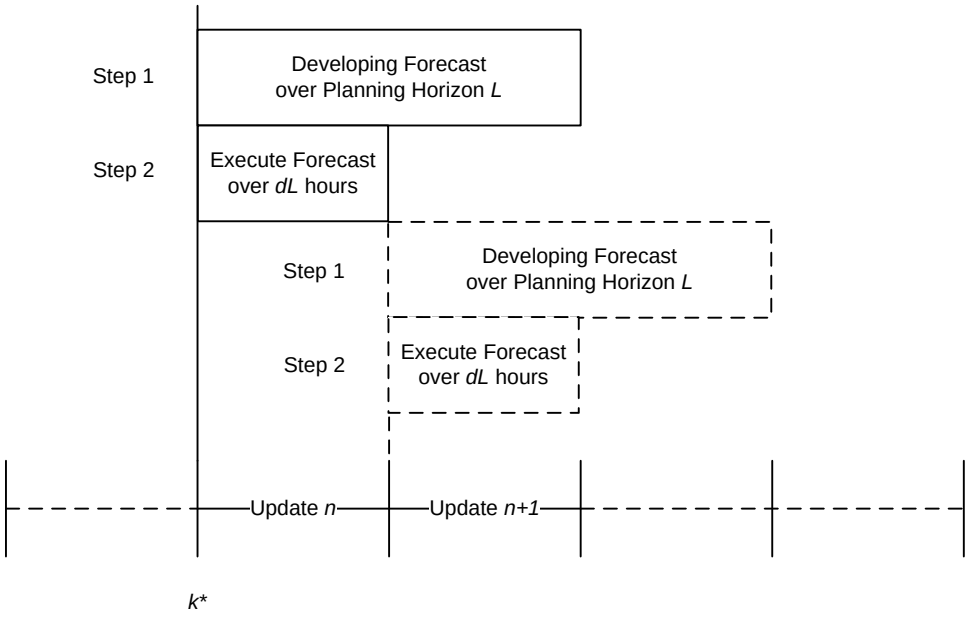

Figure 5.1: Cyclical two-stage process of planning and execution horizon.

5.3.2.3 Anchoring of the Characteristic Profile

The deviation of the measured weather variable from its forecast value at time $k*$ is described by:

$$\delta x_{k^*} = x_{k^*} - MA_{sp,k^*}(d)$$

Modifying the SPMA based on δx_{k^*} can be accomplished in a number of ways, but the underlying assumption is that the deviation will persist for the next dL hours of the execution horizon and thus the prediction will be improved. Absolute and relative δx_{k^*} modifications are considered.

Absolute Deviation Modification: At the instant in time labeled k^*, the actual measurement is compared to the forecast value. The calculated deviation δx_{k^*} is added to (or subtracted from) each element of the characteristic profile to produce the forecast for the next dL hours in the execution horizon. The basic elements described are combined to produce the dynamic short-term weather forecasting according to the strategy of absolute deviation simple prior moving average

without external forecasts:

$$MA_{asp,t}(d) = \frac{1}{d}\sum_{i=1}^{d} x_{t-\nabla i} + x_{k^*} - \frac{1}{d}\sum_{i=1}^{d} x_{k^*-\nabla i} \qquad t \in [k^*, k^* + L]$$

or,

$$MA_{asp,t}(d) = MA_{sp,t}(d) + \delta x_{k^*} \qquad t \in [k^*, k^* + L] \tag{5.4}$$

where δx_{k^*} will be positive or negative according to the k^* measured value, and $\{\hat{x}_t\}$ represents the modified vector used for dL hours in the execution horizon.

Relative Standard Deviation Modification: The absolute vertical translation according to δx_{k^*} neglects the stochastic dispersion of weather data that has occurred in the past horizon. With the inclusion of standard deviation calculations for each weather variable and each hour of the day, it is possible capture how much the measurement deviates from its mean value μ in a relative sense. For example, if the observed deviation is 3 and the standard deviation σ for the current hour of the day is 4, then the SPMA for each hour in the prediction horizon L is corrected by moving each hourly average up or down the same number of standard deviations (here 3/4). Of course, if σ were constant for all hours of the day the relative deviation modification reduces to the absolute deviation.

The modification factor for any given time k^* can be described by:

$$R_{sd} = \frac{x_{k^*} - \mu}{\sigma_{k^*}} = \frac{x_{k^*} - MA_{sp,k^*}(d)}{\sigma_{k^*}}$$

The relative deviation modified SPMA model can then be calculated as:

$$MA_{rsp,t}(d) = MA_{sp,t}(d) + R_{sd}\sigma_t(d) \tag{5.5}$$

where σ_t is the standard deviation of weather variable x at time-of-day t.

5.3.2.4 SPMA Including External Forecasts

An on-site, data-driven perspective is taken. However, the literature has shown that external forecast provided by weather services could aid in performance. Common external predictions are

typically limited to the next day's extreme temperatures and dew points. Scaling the SPMA with external forecast knowledge could yield improved results when the coming day's weather is irregular as compared to recent observations; through the use of IWEC data the extremes in weather are assumed to be perfectly predicted for the next 24 hours.

The swing or range (extreme value difference) of the forecast weather variables is described by:

$$\delta x_{sw} = x_{max,ext} - x_{min,ext}$$

Scaling of the forecast depends on the forecast swing ratio, i.e. external forecast of the swing to that predicted by the SPMA. Scaling the characteristic profile according to this strategy is accomplished as:

$$R_{sw} = \frac{\delta x_{sw}}{\max\limits_{t\in[1,24]} \left\{ \frac{1}{d} \sum_{i=1}^{d} x_{t-\nabla i} \right\} - \min\limits_{t\in[1,24]} \left\{ \frac{1}{d} \sum_{i=1}^{d} x_{t-\nabla i} \right\}}$$

Even with the swing modification, the forecast weather variable for the next hour may deviate from the measured data. Again, an adjustment for the deviation introduced by updated forecasts is accomplished by anchoring to the measured value and described by δx_{k^*}; this is calculated after the R_{sw} scaling due to distortion.

When scaling based on external forecasts is applied to the SPMA, the absolute deviation model becomes:

$$MA_{asp,e,t}(d) = R_{sw} MA_{sp,t}(d) + \delta x_{k^*} \qquad t \in [k^*, k^* + L] \tag{5.6}$$

and the relative deviation model becomes:

$$MA_{rsp,e,t}(d) = R_{sw} MA_{sp,t}(d) + R_{sd}\sigma_t(d) \qquad t \in [k^*, k^* + L] \tag{5.7}$$

At time k^*, anchoring can proceed using either the absolute or relative modifications. The distortion of the characteristic curve, both in terms of the dilation based on external swing forecasts and absolute (or relative) translation according to δx_{k^*}, can allow response to recent weather observations and improve forecast accuracy.

5.3.2.5 Exponentially Weighted Moving Average (EWMA)

Accurate forecasts may be more heavily influenced by recent observations. Seasonality enforces this concept and the chaotic, nonstationary nature of weather may further emphasize it. The SPMA relied on two months of finite past horizon for developing characteristic profiles. In an exponentially weighted moving average (EWMA), with a theoretically infinite past horizon, more recent observations have a larger impact on the characteristic profile. As the weighting tends toward zero, the older observations have a diminishing influence and at some point the infinite past horizon becomes finite for all practical purposes.

The weighting scheme is a simple exponentially (geometric) decreasing constant applied to each weather variable according to its temporal distance from t. The discount factor α acts to diminish the influence of older observations, and weights $\alpha(1-\alpha)^i$ are successively applied according to the forecast model equation:

$$MA_{ew,t} = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i x_{t-\nabla i} \qquad (0 < \alpha \leq 1) \tag{5.8}$$

The sum of all weights must be unity, i.e.,

$$\sum_{i=0}^{\infty} \alpha(1-\alpha)^i = 1$$

The choice of α is influenced by the system under consideration to dampen older observations in a suitable fashion. For the models considered in this analysis the α value was optimized during the training phase via an exhaustive search between $\alpha = 0.05$ and $\alpha = 0.95$, with $\delta\alpha = 0.05$ steps, to find the minimum CV value.

5.3.2.6 EWMA Including Modifications

The modifications applied to the SPMA, including absolute/relative anchoring and scaling based upon external forecasts, were also investigated for EWMA models. The absolute deviation model can be stated as

$$MA_{aew,t}(d) = MA_{ew,k^*}(d) + \delta x_{k^*} \tag{5.9}$$

and the relative deviation model as

$$MA_{rew,t}(d) = MA_{ew,t}(d) + R_{sd}\sigma_t(d) \tag{5.10}$$

Similarly, the expressions utilizing external weather forecasts are for the absolute deviation model

$$MA_{aew,e,t}(d) = R_{sw}MA_{ew,t}(d) + \delta x_{k^*} \tag{5.11}$$

and for the relative deviation model

$$MA_{rew,e,t}(d) = R_{sw}MA_{ew,t}(d) + R_{sd}\sigma_t(d) \tag{5.12}$$

For each of the models $t \in [k^*, k^* + L]$.

5.3.3 Neural Networks (NN)

With the ability to perform any arbitrary nonlinear mapping of input-output patterns, neural networks appear to be well suited for forecasting that exists in the *data-rich* and *theory-poor* domain, complementing the data-driven perspective. Neural networks set themselves apart from sequential computation by distributing the computational tasks of a problem onto many identical simple units ("neurons") that are highly interconnected and can work in parallel. Neural networks can approximate any continuous function to a high degree of accuracy, and are therefore well suited for time series prediction. Time series prediction can be seen as the task of finding regularities and dependencies in the data set, and neural networks can be taught to emulate the underlying dynamics of the system. Due to the unpredictable nature of noise, only the deterministic part of the problem can be predicted; the network establishes the nonlinear functionality.

A network is trained on a time series by presenting it a fixed number of previous data points $x_t, x_{t-1}, \ldots, x_{t-n}$ resulting in a fixed *time window* (or *tapped-delay line*) to predict the future values. The parameter n determines the *dimensionality* of the prediction problem. Increasing dimensionality increase performance to the point of diminishing returns, which has been formally described by *Takens' Theorem* that states no more than $m = 2df + 1$ measurements of a variable

are necessary to correctly predict the future value of this variable, where *df* is the effective number of the system's degrees of freedom [143]. However, Takens' Theorem is derived using noise-free differential equations, and trial-and-error solutions are required for realistic implementation.

Neural networks can be classified into dynamic and static categories. Static (feedforward) networks have no feedback elements and contain no delays; the output is calculated directly from the input through feedforward connections. In dynamic networks, the output depends not only on the current input to the network, but also on the current or previous inputs, outputs, or states of the network. Moreover, dynamic networks can also be divided into two categories: those that have only feedforward connections, and those that have feedback, or recurrent, connections. Dynamic networks are generally more powerful than static networks (although somewhat more difficult to train). Because dynamic networks have memory, they can be trained to learn sequential or time-varying patterns.

5.3.3.1 Focused Time Delay Neural Network (FTDNN)

The most straightforward dynamic network, which consists of a feedforward network with a tapped delay line at the input is called the focused time-delay neural network. This is part of a general class of dynamic networks, called focused networks, in which the dynamics appear only at the input layer of a static multilayer feedforward network. The architecture is equivalent to that of Figure 5.2, but devoid of the (square) exogenous input units. This network is well suited to time-series prediction and will be thus adopted and named FTDNN.

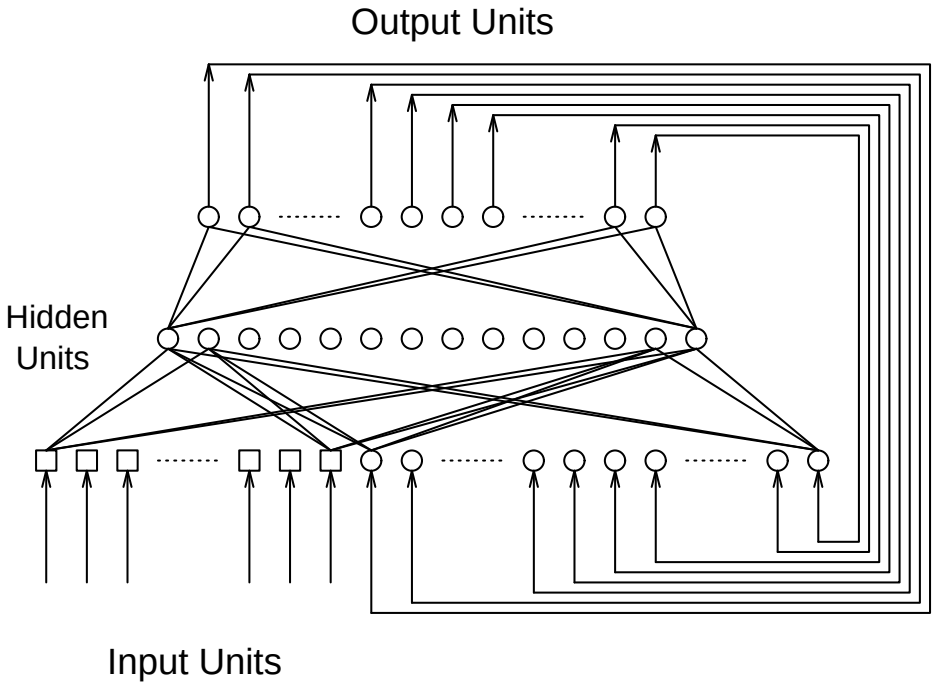

Figure 5.2: Recurrent neural network architecture.

5.3.3.2 Nonlinear Autoregressive with Exogenous Input Neural Network (NARX)

The second NN investigated is the nonlinear autoregressive with exogenous input (NARX) neural network, shown in Figure 5.2, which uses a tapped-delay line for both the recurrent output as well as clock representation (sine and cosine) of the time-of-the-day and day-of-the-year as exogenous inputs per the recommendation of Dodier and Henze [138]. The motivation behind the choice of time lags is as follows: Past values of environmental variables are relevant to prediction of current variables, so including past values of inputs should improve the accuracy of predictions. However, it is clear that there are limits to the useful time lag. If the time lag is very long (a month, say) the time-lagged input will be almost completely uncorrelated with the current variable. If, on the other hand, the time lag is very short (a minute, say) the time-lagged input will be almost completely redundant with the current input variable, providing no new information. To choose an

appropriate intermediate time lag the autocovariance function was inspected:

$$C(\tau) = \frac{1}{N}\sum_{t}(x_t - \bar{x})(x_{t-\tau} - \bar{x}) \tag{5.13}$$

for each input variable at multiples of one hour. The first zero of the autocovariance represents a time at which the lagged input is uncorrelated with the current value, i.e. the two points are statistically independent, so it is the shortest time lag which brings in the most new information.

Optimization: For both the FTDNN and the NARX models, the tapped line length was bracketed around the first zero of the autocovariance function of the dry-bulb temperature T_{db} of 11 hours as reported by Dodier and Henze [138]. The optimal network architecture was found by exhaustively searching in the range of 9 to 13 hours for the lag and 20 to 30 neurons for the size of hidden layer until a minimum CV was found in training. Limiting the optimization of network architecture to the stated ranges allowed a balance between the extremely long training times, the large study, and practical implementation. It should be noted that the Levenberg-Marquardt training algorithm was adopted for the FTDNN architecture and Bayesian regularization for the NARX architecture. In addition, early stopping and data scaling have been utilized in both neural networks.

5.4 Results and Discussion

The 14 different forecasting models of Table 5.2 were programmed in a commercially available technical computing environment. For each of the 11 geographic locations in Table 5.1, the four weather variables (I_{gh}, T_{db}, T_{wb}, ϕ) were separately predicted and the performance was described in terms of the metrics of Equations 5.1 and 5.2; the evaluation was accomplished separately for each of the four seasons. Nearly 2,500 prediction tasks were conducted to ascertain the required complexity of short-term weather forecasting model for MPC.

A general impression of each forecasting model's performance was obtained by collating the metrics for an exclusive prediction task, e.g. relative humidity, and the expected circumstances, i.e. yearly basis. Therefore, for each forecasting model's performances, the mean was taken across all

geographic locations and all seasonal performances. Given in percentage CV and MBE, Tables 5.3 and 5.4 present the results, respectively. Entries labeled 'na' indicate particularly poor prediction performance, exceeding 100% CV or MBE such that the metrics lose their interpretation other than signaling inferior performance. Observing Tables 5.3 and 5.4, it may appear the magnitude of the values is large. However, a formidable challenge of the $dL = 6$hour execution of prediction was intentionally put forth to thoroughly examine the forecasting models' performance; it should be kept in mind the values are calculated by concatenating all 6-hour predictions for the defined 'exclusive prediction task'. Furthermore, the largest metrics are associated with I_{gh} and T_{wb}, which have a diurnal nature exhibiting more sporadic behavior than T_{db} or ϕ.

Table 5.3: Forecasting Models' Mean %CV over all Locations for each Weather Variable Predicted.

Model	Igh	Tdb	Twb	RH
1	na	na	na	27
2	69	na	na	90
3	91	19	55	12
4	78	19	55	12
5	57	20	92	13
6	83	20	95	13
7	61	39	na	21
8	57	60	na	65
9	49	20	62	11
10	69	21	62	12
11	45	22	92	12
12	67	22	94	12
13	na	78	na	19
14	na	31	93	10

Table 5.4: Forecasting Models' Mean %MBE over all Locations for each Weather Variable Predicted.

Model	Igh	Tdb	Twb	RH
1	74	na	na	22
2	35	na	na	76
3	53	13	35	8
4	35	12	36	8
5	32	13	59	8
6	42	13	61	8
7	30	29	na	16
8	24	45	na	49
9	24	13	40	7
10	28	13	40	7
11	21	14	62	8
12	29	14	63	8
13	na	17	50	10
14	56	22	68	7

5.4.1 Performance Comparison

It was necessary to analyze the performance (metric) variance in order to compare forecasting models and make recommendations for application in distinct prediction tasks. This investigation of variance is only possible within a given geographic location, i.e. a unique "data source", and furthermore it was assumed the model is used for predicting one weather variable throughout the entire year. Therefore, changes in seasonal performance of a given model are considered a nuisance factor effect, i.e. a background effect that needs to be considered but is not the primary interest. The difference in forecasting performance of the 14 models is the concern and not the difference in an individual model's season-to-season forecasting performance, but the latter must still be considered.

As described in Hollander and Wolfe [144], the Friedman test is a nonparametric randomized block design. It was used to compare the performance effects of the 14 models while blocking the nuisance seasonal effects. The two-way analysis of variance (ANOVA) was initially considered, but the models' residuals were not normally distributed with equal treatment (forecasting model) variance as required by ANOVA assumptions. Friedman's test is a rank sum procedure that relies on the comparison of parameter location (median), allowing determination of the models' relative forecasting performance. The test assumes the data comes from a continuous population distribution (satisfied by IWEC weather data) and all observations are mutually independent (satisfied by programming independence). The null hypothesis states that apart from minor nuisance block effects, the parameter location is equal for each treatment, i.e. there is no difference in forecasting model performance. The alternative hypothesis is that a significant difference exists among the treatments.

The Friedman statistic can be calculated as:

$$Fr = \left[\frac{12}{mn(n+1)} \sum_{j=1}^{n} R_{\bullet j}^2 \right] - 3m(n+1) \tag{5.14}$$

where separately within m independent blocks (seasons) the n treatments (forecasting models) are ranked. The treatment with the best performance is assigned rank 1 and the worst performance

rank n; there is an adjustment for ties, but it was not a concern here. $R_{i,j}$ is the rank for the i-th ($i \in 1, 2, \ldots, m$) block for the j-th ($j \in 1, 2, \ldots, n$) treatment. Within a given block, there are $n!$ possible ranking arrangements, leading to $m(n!)$ ranking possibilities when considering m blocks. If the null hypothesis is valid then each mean ranking of a treatment is equally likely. The test statistic Fr is approximately a chi-square distribution with $n - 1$ degrees of freedom. In an upper-tailed critical value test, the Fr statistic is compared to the respective χ^2 at an α level of significance when deciding to accept or reject the null hypothesis; a common value of $\alpha = 0.05$ was selected. Statistically significant differences in (yearly) forecasting performance were found, using the Friedman test and the stated level of significance, for nearly every combination of weather variable and geographic location.

With those combinations found to have statistically significant differences with Friedman's test, a (post hoc) multiple comparison technique was used to determine a specific model's superiority over another. Hochberg and Tamhane [145] recommend the following procedure as the most powerful for all pairwise comparisons:

$$\left|\bar{R}_i - \bar{R}_{i'}\right| > \frac{Q^{(\alpha)}_{n,df}}{\sqrt{2}} \sqrt{\frac{n(n+1)}{6m}} \qquad (1 \leq i \leq i' \leq n) \tag{5.15}$$

When the absolute difference of any two mean ranks $\overline{R_i}$ and $\overline{R_{i'}}$ exceeds a critical value they are deemed statistically different; $Q^{(\alpha)}_{n,df}$ is the studentized range statistic obtained from tables or computational environments. The method relies on all pairwise difference having the same variance, i.e. *pairwise balanced*, and then $(1 - \alpha)$ confidence intervals for all comparison can be calculated. The method is an extension of Tukey's honestly significant difference (HSD) method.

The multiple comparison procedure allowed determination of the forecasting models' statistically significant superiority or inferiority for the yearly prediction of a specific weather variable within a given geographic location. For example, the EWMA with absolute anchoring (model 9) was found to be the best CV performer for Atlanta dry-bulb temperature prediction, and for the same prediction task is (statistically significant) superior to model 2. It is possible that one model shows statistically significant behavior over another, yet still not have the lowest CV or MBE.

Figures 5.3 and 5.4 present the frequency of statistically significant superiority or inferiority of each model for, respectively, CV and MBE metrics. The frequencies are found by summing the number of significant events (behavior) over all geographic locations. This complements the mean CV and MBE tables, following discussions, and compactly summarizes the results of the nearly 2,500 prediction tasks.

5.4.2 Discussion

To shed light on the results, Table 5.2 is used to distinguish among forecasting models, Tables 5.3 and 5.4 provide the mean CV and MBE metric performance, and multiple comparisons results are summarized in Figures 5.3 and 5.4. For instance, models 9 and 11 differ only in the use of external (extreme) weather forecasts in the latter. Using the scaling factor R_{sw} actually causes a respective 30% and 22% mean increase in CV and MBE for T_{wb} prediction. The corresponding frequency of superiority investigation shows that the model's superiority drops from 6 to 0 for CV and 5 to 0 for MBE. Thus, scaling EWMA forecasts based on external wet-bulb temperature forecast extremes is generally not recommended, even with perfect prediction.

The short-term forecasting of global horizontal radiation I_{gh} is a difficult task with the best predictor achieving a mean CV value of 45% and MBE of 21%. The best three predictor models 11, 9, and 8 are all EWMA models, with external forecasts (11, 8) and absolute deviation modification (11, 9) and no anchoring (8). The fact that an EWMA without anchoring to the current observed value fares this well comes as a surprise when compared to the other prediction variables. In Figures 5.3 and 5.4 it appears that EWMA models provide consistently superior performance. It is interesting to note the FTDNN (13) is consistently inferior. The neural networks do not seem to provide adept pattern matching for the complex, sporadic behavior of I_{gh} and are not recommended without exogenous inputs.

The prediction of dry-bulb ambient temperature is without a doubt the most important forecasting task. The best prediction performances, with mean CV values on the order of 20% (13% MBE), were achieved by SPMA models 4, 3, 6, and 5. The EWMA models 9, 10, 11, and 12

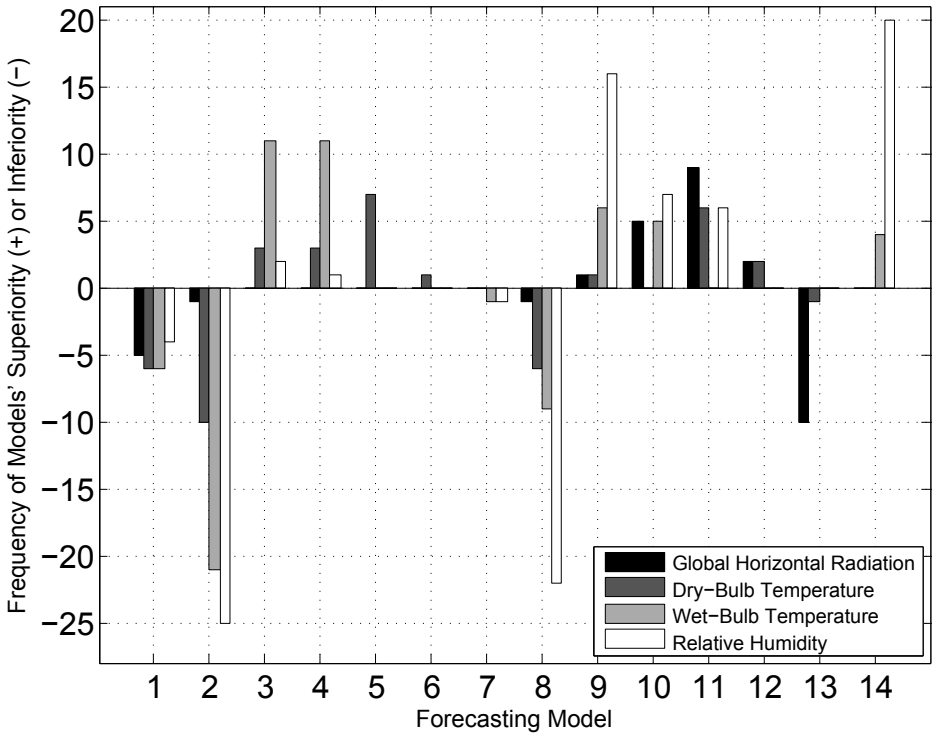

Figure 5.3: Multiple comparisons using CV.

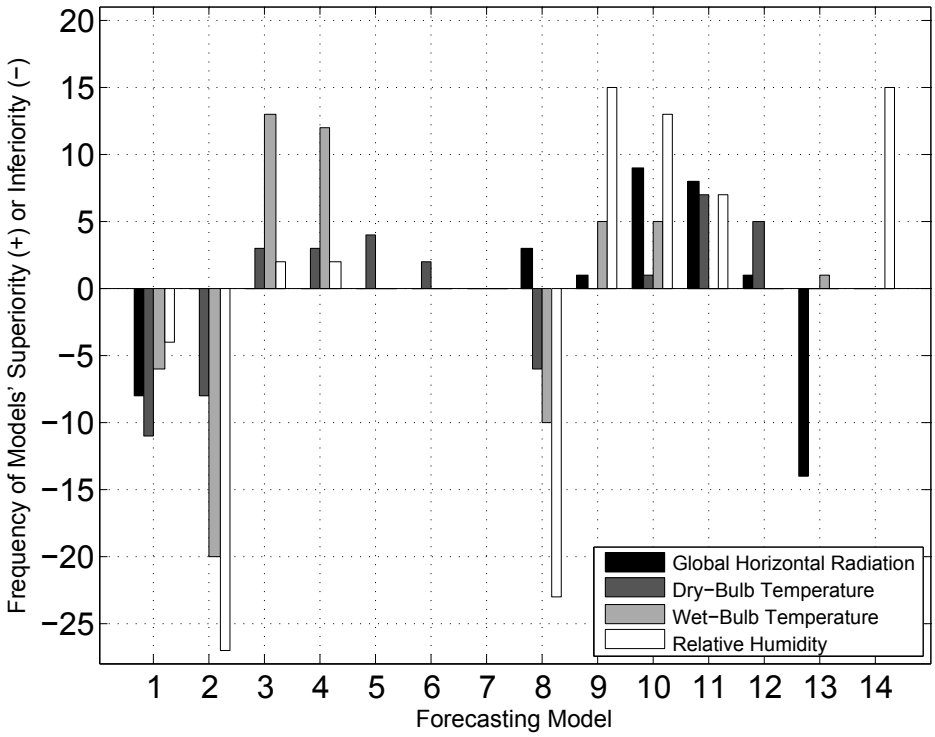

Figure 5.4: Multiple comparisons using MBE.

offered only slightly inferior performance of about 22% (14% MBE), while the complex NARX (14) neural network model achieved only 31% CV and 22% MBE. Anchoring, both absolute and relative, improves prediction performance and is highly recommended for temperature prediction. The availability of an external forecast of the next day's extreme values offers insignificant performance benefits. The FTDNN again fared very poorly.

With the increase in magnitude of CV and MBE values, it becomes apparent that the prediction of wet-bulb temperature alone is an extremely difficult task with none of the models achieving a mean CV of less than 50% or 35% mean MBE. The frequency of superiority is higher with the more complex models, but the efforts are futile when observing performance gains. In short, the use of T_{wb} in determining ambient humidity levels is discouraged, and instead, the use of relative humidity ϕ is recommended. This is due to the dependence relative humidity has on dry-bulb temperature: the prediction task is simplifed when the variables assume a fairly anticipated shape, rather than the prediction task being a completely stochastic pursuit; the argument extends to humidity ratio predictions.

Relative humidity emerged as the easiest prediction task with the best model (NARX model 14) achieving a mean CV value of 10% and MBE of 7%. However, a wide range of models provides only slightly inferior performance: EWMA models 9, 10, 11, 12 offer acceptable performance followed by SPMA models 4, 3, 5, and 6. The frequency of superiority enforces these results. Neither the FTDNN nor MA models without anchoring are recommended for the prediction of relative humidity. The inferiority is particularly high for models 2 and 8.

Example Weather Predictions: Time series plots are shown in Figures 5.5 through 5.7 to illustrate the closed-loop prediction performance over the $dL = 6$ hour prediction horizon for increasingly stochastic weather phenomena. A nearly sinusoidal dry-bulb temperature prediction, with mild trend, is offered as an example in Figure 5.5 for the location of Stuttgart, Germany during the summer. For this case, model 4 (SPMA with relative anchoring) achieves $CV = 8.6\%$ and $MBE = 5.8\%$, while model 14 (NARX) yields a $CV = 11.2\%$ and $MBE = 8.0\%$. Simple SPMA models perform as well as, or better than, the much more complex NARX model because

the daily profiles are close to the characteristic profile and the trend is accounted for through the anchoring process.

An example of relative humidity prediction in Phoenix, Arizona during the summer season is shown in Figure 5.6 as a more complex time series. The daily maximum value changes drastically from 50 to 90% and back from 90 to 30%, and oscillatory behavior on days 2 and 3 complicates the prediction task. For this case, the best model is the NARX neural network model 14 which yields $CV = 14.9\%$ and $MBE = 11.0\%$, followed by the EWMA model with absolute anchoring (model 9) achieving $CV = 19.8\%$ and $MBE = 11.0\%$.

A highly complex prediction task, the forecasting of wet-bulb temperature in Beijing, China during the spring season is presented in Figure 5.7. The best performer is the NARX neural network with a very poor performance of $CV = 74.6\%$ and $MBE = 51.1\%$. Interestingly, the much simpler SPMA model 3 follows the neural network with metrics $CV = 78.0\%$ and $MBE = 44.6\%$ as a close second.

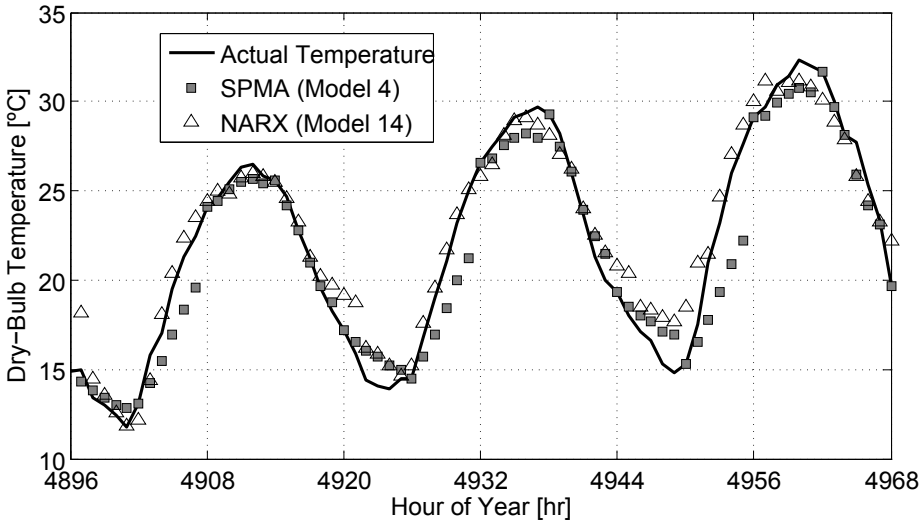

Figure 5.5: Forecast prediction in the face of simple, nearly sinusoidal weather variation.

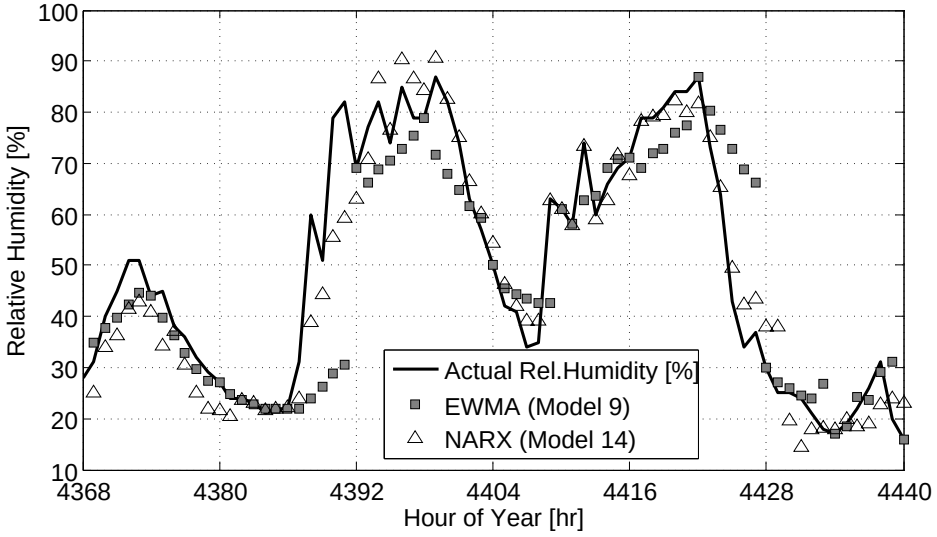

Figure 5.6: Forecast prediction in the face of complex day-to-day variations in weather.

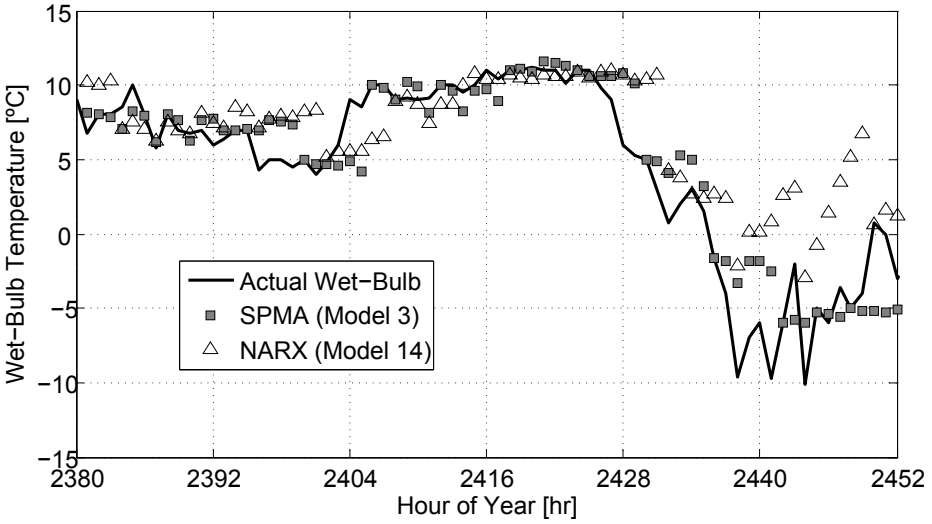

Figure 5.7: Forecast prediction in the face of highly uncertain, erratic weather variation.

5.5 Conclusions

The performance benefits of using the much more complicated focused time delay neural network (FTDNN) or the nonlinear autoregressive neural network with exogenous input (NARX) does not appear to warrant the additional efforts in predictor model development and training.

Simple prior moving average models (SPMA) are superior in predicting dry- and wet-bulb temperatures, whereas the exponentially weighted moving average models show performance gains in predicting solar radiation and relative humidity. The two methods would be recommended for the respective prediction tasks. However, for all practical purposes, the differences in performance between SPMA and EWMA are small. This important conclusion can only be gleaned from this analysis when it is ensured that forecasts can be updated at every instance dL and anchored to the known observed value of the prediction variable. Furthermore, the use of MPC in buildings with a data-driven perspective should rely on variables that vary diurnally and thus reduce the prediction task difficulty, i.e., favor the prediction of relative humidity over wet-bulb temperature, as even complex neural networks cannot establish the underlying dynamics solely from data. As such, the use of simple time series analysis within the cyclical two-stage model predictive control process of policy planning followed by execution outperforms even the most complicated nonlinear autoregressive neural with exogenous input (NARX) model.

Chapter 6

An Energy Signal Tool for Decision Support in Building Energy Systems

6.1 Introduction and Motivation

Stakeholders of assorted interests are increasingly concerned with the energy performance of the built environment.[1] Increasing commitment to energy efficiency, cost-minimal retrofits, and renewable energy integration has coincided with the availability of commercial and open-source building energy simulation engines. Model-based approaches have become the norm, with engineering design accelerating its reliance on software. It is hypothesized that, beyond building design applications, model-based engineering of buildings can be extended to encompass a building's multi-decade life cycle. Of particular interest is the operational energy performance, where tradeoffs in comfort and energy consumption can be hidden, and the establishment of "normal behavior," as distinguished from "faulted behavior," is nontrivial. Research interests lie in data-driven models for decision-making processes that are flexible, adaptable, and can evolve with the engineered system.

A balance must be struck between model sophistication and available data. One may have scores of utility bill data available but little understanding of an appropriate, physically relevant model. Or one might have an exceedingly detailed physical model available but its real-world validity is still questionable because calibration has been performed against sparse utility data. Pattern recognition or classification can be used to ascertain the validity of a model and the

[1] Chapter 6 largely derives from Henze et al. [146] and authorship is attributed approximately one-third to Anthony R. Florita for fault modeling and Bayesian Analysis; the first author contributed decision analysis content, the second author gray-box modeling content, and the last two author contributed editing.

value of data; however, building applications are in their infancy. At the building systems level, monitoring-based heating, ventilation, and air conditioning (HVAC) commissioning [147] and chiller fault detection [148] have shown promise.

The interpretation of patterns might be further aided by providing real-time, appliance-level power management and occupant feedback for sociotechnical energy conservation [149]. At the whole-building level, participation in the smart grid via approaches such as energy storage may entail value-cognizant electricity demand shifting and shaping [112]; the value to the building owner is likely different than that to the electricity grid. Data mining and knowledge discovery tasks have the ultimate goal of predictive diagnostics for buildings and their systems, and have shown acceptable levels of misclassification in the face of the evolving, nonstationary behavior common to buildings [150]. Sector-wide studies include modeling the evolution and refurbishment of the German heating market (for 2050 goals) and its impact on carbon emissions [151].

The goal of the energy signal tool research is to enable owners and operators of commercial buildings to quickly (in a matter of seconds) attain insight into how their buildings' energy use compares against the likely range of expected energy consumption over a given time period (days, weeks, months, or years). The output of the energy signal tool is a simple traffic light indicator that summarizes energy consumption relative to model-based expectations. To find the appropriate value of the indicator, the energy signal tool carries out an analysis of building energy use, taking uncertainty and misclassification costs into account. As illustrated in Figure 6.1, the energy signal tool process begins with an operational energy model of a building to provide expected energy performance, but recognizes that any model only approximates reality.

Previous research explored how gray-box models are obtained and calibrated from noisy data [152], and results are extended here to include HVAC systems. The term operational derives from the desire to consider only a few influential variables within the model and to use them in real-time applications while learning from data as they are gathered. The simplified operational models are sufficient when coupled to uncertainty analysis and misclassification costs of relatively simple building types such as big box retail. Work is currently underway to develop an open-source tool

based on the OpenStudio development effort that would allow the decision analysis to be applied to arbitrarily complex multi-zone buildings.

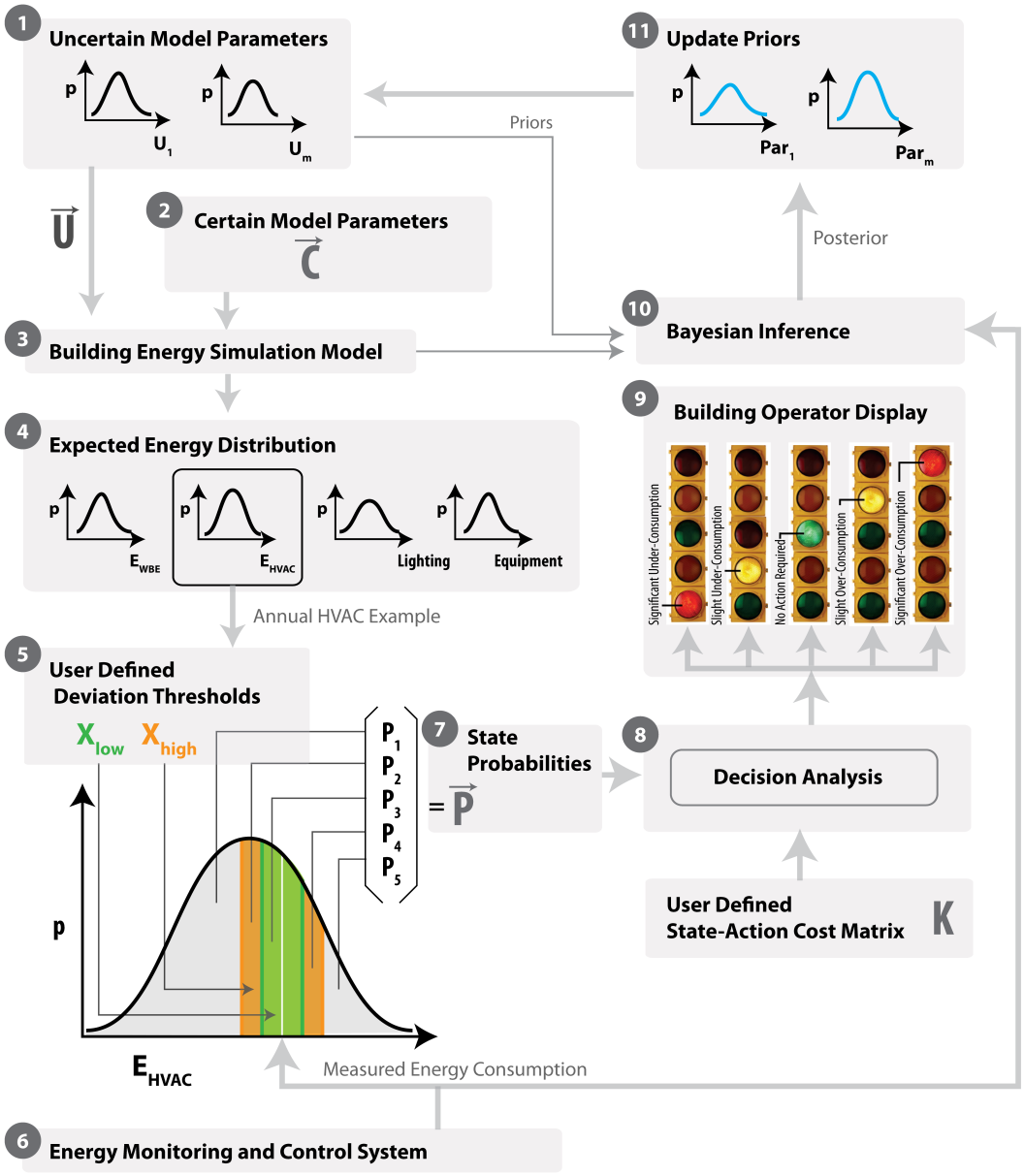

Figure 6.1: Energy signal tool flowchart.

A Bayesian probabilistic approach was adopted here to update beliefs about uncertainties in light of new data. Over time, the energy signal tool learns improved assumptions for input parameter uncertainties by incorporating measured building data into a Bayesian inference process.

Unobserved variables are inferred from data and physical modeling. The range of all possible values is divided into five exhaustive and mutually exclusive intervals, labeled 1–5 in the figure, which represent predicted energy use that is substantially lower, somewhat lower, more or less the same, somewhat higher, and substantially higher than observed. The probability that energy use (at either the whole-building or the end-use level) falls into a given range of values, is computed as the integral of the energy use probability distribution over that interval. User-defined thresholds determine the tool's sensitivity and are driven by the operator's risk appetite. Then applied utility theory is applied to find the most appropriate action given an assumed cost of misclassification of each action (i.e., each traffic light color) in each state (i.e., each energy use interval probability). The expected cost of misclassification is the cost matrix multiplied by the probability vector. The element of the expected cost that has the lowest value is chosen.

To illustrate the operation of the energy signal tool, examples are given of its output in various energy use scenarios and review Bayesian updates to model parameters.

6.1.1 Nomenclature

$\vec{a}$	action vector
a_{opt}	optimal action
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$	state space model matrices
C	opaque building shell thermal capacitance
C_c	roof thermal capacitance
$C_{c,1}$	external node roof thermal capacitance
$C_{c,2}$	internal node roof thermal capacitance
C_e	exterior wall thermal capacitance
$C_{e,1}$	external node exterior wall thermal capacitance
$C_{e,2}$	internal node exterior wall thermal capacitance
C_f	floor thermal capacitance
$C_{f,1}$	external node floor thermal capacitance
$C_{f,2}$	internal node floor thermal capacitance
C_i	internal thermal capacitance
$C_{i,1}$	node 1 internal thermal capacitance
$C_{i,2}$	node 2 internal thermal capacitance
C_p	air thermal capacitance
C_z	zone thermal capacitance
CM	zone air capacitance multiplier
COP	coefficient of performance
D	measured data
DX	direct expansion
e_k	transfer function heat gain history coefficient
$E(x)$	expected value of x
E_{meas}	measured predicted energy consumption
E_{mod}	model-predicted energy consumption

$E_{0,low}$	lower threshold of low deviation
$E_{1,low}$	upper threshold of low deviation
$E_{0,high}$	lower threshold of high deviation
$E_{1,high}$	upper threshold of high deviation
EQP	equipment
G	central green light action
GM	internal gain multiplier
h_{fg}	heat of vaporization of water
HVAC	heating, ventilation, and air conditioning
i	action index
j	state index
K	cost matrix (decision analysis) or knowledge (inference)
LTG	lighting
m_{air}	mass of air in the zone
m	transfer function heat gain history order
M	model output
$\dot{m}_{inf}$	infiltration mass flow rate
$\dot{m}_{SA}$	supply air mass flow rate
ML	much lower state
MH	much higher state
NSU	nighttime setup
n	number of past inputs
$\vec{P}$	state probability vector
p	probability
$q_{occ,lat}$	occupant latent gains
$\dot{Q}_{ep}$	surrogate or measured sensible zone load
$\dot{Q}_{g,c}$	convective portion of internal gains (lighting, occupants, and equipment)

$\dot{Q}_{g,r,c}$	radiative fraction of internal gains applied to ceiling surface
$\dot{Q}_{g,r,e}$	radiative fraction of internal gains applied to vertical wall surface
$\dot{Q}_{g,r+sol,w}$	radiative portion of internal gains and solar radiation through glazing
$\dot{Q}_{inf}$	infiltration heat gain
$\dot{Q}_{sol,c}$	solar radiation transmitted through opaque ceiling/roof surfaces
$\dot{Q}_{sol,e}$	solar radiation transmitted through opaque vertical exterior surfaces
$\dot{Q}_{sol,w}$	solar radiation transmitted through glazing
$\dot{Q}_{sh}$	sensible convective heat gain to zone air
$\dot{Q}_{zs}$	sensible zone load
$\dot{Q}_{rom}$	reduced-order model predicted sensible zone load
r	number of elements in the input vector $\mathbf{u}$
R_1	combined heat transfer coeff. to opaque shell mass node
R_2	cond. coeff. between mass and internal surface node
R_3	conv./rad. coeff. b/w surf. and zone air temp. nodes
R_c	roof thermal resistance
$R_{c,1}$	roof combined external convection and radiation coefficient
$R_{c,2}$	roof conduction resistance
$R_{c,3}$	roof internal combined convection and radiation coefficient
R_e	exterior wall thermal resistance
$R_{e,1}$	combined external convection and radiation coefficient
$R_{e,2}$	exterior wall conduction resistance
$R_{e,3}$	exterior wall internal combined convection and radiation coefficient
R_f	floor thermal resistance
$R_{f,1}$	ground conduction coefficient
$R_{f,2}$	floor conduction resistance
$R_{f,3}$	floor internal combined convection and radiation coefficient
R_i	internal partition thermal resistance

$R_{i,1}$	internal partition combined convection and radiation coefficient
$R_{i,2}$	internal partition conduction resistance
$R_{i,3}$	internal partition combined convection and radiation coefficient
R_w	glazing thermal resistance
RA	return air
RH	upper red light action
RL	lower red light action
ROM	reduced-order model
RTU	rooftop unit
S	similar state
SA	supply air
SH	somewhat higher state
SHGC	solar heat gain coefficient
SL	somewhat lower state
$\mathbf{S}$	transfer function input coefficient matrix
t	time or time index
T_a	outdoor air temperature
T_c	ceiling node temperature
$T_{c,1}$	external roof node temperature
$T_{c,2}$	internal roof node temperature
T_e	exterior wall node temperature
$T_{e,1}$	external exterior wall node temperature
$T_{e,2}$	internal exterior wall node temperature
T_f	floor node temperature
$T_{f,1}$	external floor node temperature
$T_{f,2}$	internal floor node temperature
T_g	ground temperature

T_i	internal partition temperature
T_m	opaque building shell thermal temperature
T_s	(pseudo) internal surface temperature
T_z	zone air temperature
$\bar{T}_z$	average zone air temperature over timestep
$\mathbf{u}$	input variable vector
WBE	whole building electricity consumption
W_z	zone air humidity ratio
W_{OA}	outdoor air air humidity ratio
W_{SA}	supply air humidity ratio
$\mathbf{x}$	state variable vector
$\dot{\mathbf{x}}$	state variable first derivative vector
X_{high}	definition of high level of deviation threshold
X_{low}	definition of low level of deviation threshold
$\mathbf{y}$	output variable vector
YH	upper yellow light action
YL	lower yellow light action
σ_ϵ	measurement noise
$\Delta\tau$	time step

6.2 Literature Review

Most uncertainties in building energy performance are addressed during the design phase. The evolution of a given design involves a sequence of decisions by various domain experts and has implications in thermal, visual, and acoustical performance [153]. Competing objectives such as energy consumption, environmental performance, and financial costs warrant multi-objective optimization for decision-making [154, 155]. Although engineering tradeoffs lead to numerous optimal and near-optimal solutions; e.g., Pareto fronts [156], early design choices lead to the building's

ultimate sustainability [157]. Confounding the problem is information sharing with conflicting objectives in the collaborative design process [158]. However, the primary, uncertain drivers in the design process include (1) (micro)climate variables [159], which may not be appropriately captured by typical meteorological data; (2) occupancy patterns and dynamics, which may be hard to capture with traditional diversity factor approaches [160]; and (3) consideration for the existing infrastructure where the building will be constructed, which may be far from ideal [161]. Judkoff et al. [162] described the sources of difference between simulation and reality. Recent interest lies in sustainable designs with renewable energy systems [163], net zero energy buildings [164], and overall healthy and productive buildings [165, 166].

Energy management or measurement and verification within existing building energy systems must face a plethora of uncertainties, including (but not limited to) noisy sensors, point measures of distributed phenomena (e.g., air temperature), and unobserved variables. To capture complex, nonlinear, and multivariable interactions, mathematical approaches such as Gaussian processes [167, 168], multi-agent decision-making control strategies [169], and Bayesian-calibrated energy models [16, 170] have been used. Furthermore, with the proliferation of wireless sensor networks in smart buildings [171], interest in assessing performance has extended beyond energy into mold growth and remediation [172], as well as disaster preparedness and management [173, 174] for events such as fires [175], earthquakes [176], and bioterrorist attacks [177]. The literature shows that the need for decision support within operational building settings is vast, yet a balance between risk and situational usefulness needs to be attained.

Many authors have devised frameworks for decision support in various building energy performance settings. Augenbroe et al. [178] described a tool with an investment strategy for energy performance decision-making for existing buildings with viable refurbishments via optimization. Kolokotsa et al. [179] analyzed and categorized buildings for specific actions or groups of plans in a methodology for decision support of building energy efficiency and environmental quality, including real-time operation and offline decision-making. Das et al. [180] considered building maintainability using an analytical hierarchy process to balance budget requirements with performance standards

for nine building systems, including input from 37 facilities management experts. Gultekin et al. [181] developed a decision support system for guidance in "green retrofits" to identify key criteria and feasible alternatives. Mohseni et al. [182] offered a comprehensive decision-making methodology for condition monitoring to guide building asset managers, aiding capital investments and expenditures. In a series of papers, Lee et al. [183, 184, 185] detailed process models for decision support in energy-efficient building projects, and campus-scale infrastructures, and summarized a "workbench" for uncertainty quantification, respectively. Collectively, i.e., taking this series of three papers together, a decision support framework was provided.

6.3 Methodology

6.3.1 Modeling Environment

For prototyping the energy signal tool, the simulation study required the validation of the operational building energy model as detailed in the following sections. In practice, a measurement campaign combined with system identification techniques would be required before the energy signal tool is implemented. Because of its Bayesian learning approach, the process could be automated with a basic knowledge of the model's structure.

6.3.1.1 Retail Building Simulation Models

The U.S. Department of Energy's EnergyPlus standalone retail reference building [186, 187], post-1980 construction, was used as a relatively simple first application for prototyping and testing the energy signal tool. An isometric view of the original five-zone retail building is shown in Figure 6.2, along with a plan view of model zoning shown in Figure 6.3. One zone is dedicated to the entry vestibule, two slender zones to the left and right of the vestibule have glazing and are assumed to be affected by solar gains, a very large core retail zone occupies about 90% of the floor area. Finally, a loading and storage zone covers the back of the store. Selected model details are highlighted in Table 6.1. This five-zone EnergyPlus model was used to generate simulated

operational data for use in developing the reduced-order building energy models described in the following subsections. Surrogate data were preferred here over real measurements so that latent variables could be controlled in the experimental study.

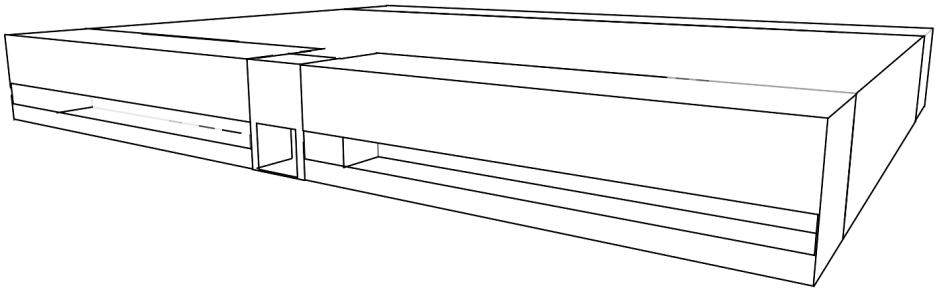

Figure 6.2: Isometric view of five-zone retail building model.

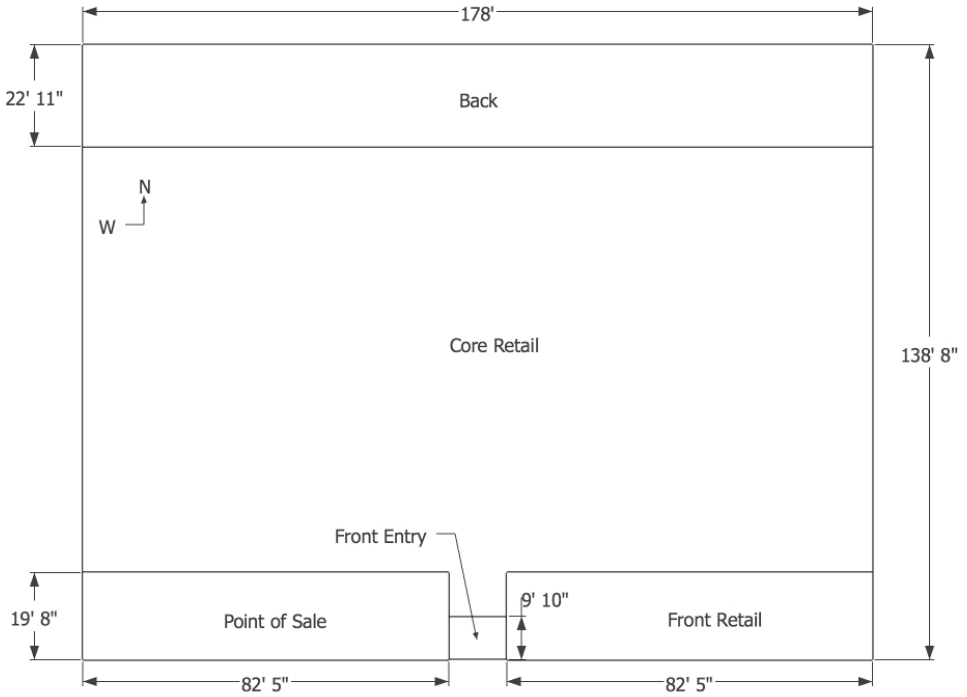

Figure 6.3: Zone plan of five-zone retail building model.

Table 6.1: Selected EnergyPlus Model Details.

Property	Value	Units
Vintage	1980	year
Volume	13 984	m^3
Conditioned floor area	2294	m^2
Bldg. avg. U-value (no film, excluding floor)	0.418	$W\,m^{-2}\,K^{-1}$
Ext. wall U-value (no film)	0.621	$W\,m^{-2}\,K^{-1}$
Roof U-value (no film)	0.314	$W\,m^{-2}\,K^{-1}$
Floor U-value (no film)	12.904	$W\,m^{-2}\,K^{-1}$
Internal thermal capacitance	450	$MJ\,K^{-1}$
Internal thermal capacitance per floor area	196.2	$kJ\,K^{-1}\,m^{-2}$
Infiltration	1.01	ACH
Glazing fraction	7	%
Glazing U-factor	3.354	$W\,m^{-2}\,K^{-1}$
Glazing solar heat gain coefficient	0.385	fraction
Lighting power density	32.3	$W\,m^{-2}$
Equipment power density	5.23	$W\,m^{-2}$
Occupant density	7.11	m^2/person
HVAC system	CV-DX	-

6.3.1.2 Inverse Gray-Box Building Model for Operational Applications

The inverse gray-box modeling approach developed for this work is largely based on methods described by Braun and Chaturvedi [188, 189]. For the application presented in this work, it is important to be able to predict transient cooling and heating requirements for the building using inverse models that are trained using on-site data. Inverse models for transient building loads range from purely empirical or "black-box" models to purely physical or "white-box" models. Generally,

black-box (e.g., neural network) models require a significant amount of training data and may not always reflect the actual physical behavior, whereas white-box (e.g., finite difference) models require specification of many physical parameters. Braun and Chaturvedi introduced a hybrid or "gray-box" modeling approach that uses a transfer function with parameters that are constrained to satisfy a simple physical representation for energy flows in the building structure. A robust method was also presented for training parameters of the constrained model, wherein (1) initial values of bounds on physical parameters are estimated from a rough building description; (2) better estimates are obtained using a global direct search algorithm; and (3) optimal parameters are identified using a nonlinear regression algorithm. They found that 1 to 2 weeks of data are sufficient to train a model so that it can accurately predict transient cooling or heating requirements.

Previous to the work by [188, 189], [190, 191, 192, 193, 194], developed a modeling scheme consisting of several lumped parameters with direct correspondence to reality and correspondence to a detailed model. The model, used in combination with field data, enabled empirical determination of the input parameters, thereby reducing model uncertainty.

Inverse gray-box models may be based on the approximation of heat transfer mechanisms by an analogous electrical lumped resistance-capacitance network. This approximation creates a flexible structure that allows the modeler to choose the appropriate level of abstraction. Model complexity can range from representing entire systems with a few parameters to modeling each heat transfer surface with numerous parameters. Depending on the model structure and complexity, parameters can approximate the physical characteristics of the system. Model parameters are then identified through a training period with measured data.

Figure 6.4 shows the 21-parameter thermal network representations that Braun and Chaturvedi [188]; Chaturvedi et al. [189] found to work well. Other forms have been considered in this work and are described below. A separate 3R2C network is used to represent external wall, ceiling, ground, and internal wall heat transfer. Looking at the 3R2C network for exterior walls, for example, $R_{e,1}$ could be thought to represent a combined external convection and radiation coefficient, $R_{e,2}$ wall conduction resistance, and $R_{e,3}$ internal combined convection and radiation coefficient to

the zone air node. Solar gains from opaque elements are represented by $\dot{Q}_{sol,e}$ applied to the external surface nodes (e.g., $T_{e,1}$ and $T_{c,1}$). Storage is neglected for glazing elements that are represented by a single resistance R_w. Solar gains directly entering the zone through glazing are distributed among internal partition nodes $T_{i,1}$ and $T_{i,2}$ as $\dot{Q}_{sol,w}$. Internal gains are split into convective and radiant fractions. Convective fractions are applied directly to the zone air node T_z as $\dot{Q}_{g,c}$. Radiant portions are applied to interior surface nodes $T_{e,2}$ and $T_{c,2}$ as $\dot{Q}_{g,r,e}$ and $\dot{Q}_{g,r,c}$, respectively. (Split is proportional by surface area.)

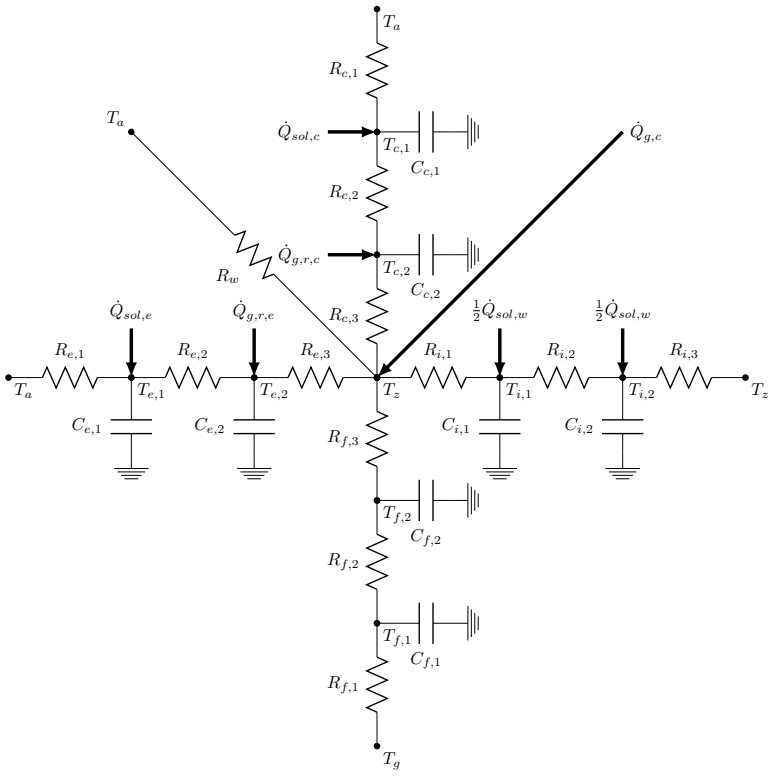

Figure 6.4: Twenty-one-parameter thermal RC network.

When using the inverse gray-box modeling approach described in this work, questions naturally arise about the RC network structure that is most appropriate for the modeling task. Selecting a very complex model structure results in a difficult parameter estimation task, but too simple a model may not appropriately capture the desired dynamics. In this research the reduced-order modeling (ROM) environment was developed to allow for model structure flexibility, so this ques-

tion may be investigated. As previously mentioned, various RC network forms have been considered in this work ranging from five to 21 parameters. Because the 21-parameter model was previously introduced, discussion will begin with the 18-parameter model (see Figure 6.5). This model can be considered a subset of the 21-parameter model with the internal surface heat transfer elements simplified to 1R1C. This reduced the parameter estimation procedure three parameters and kept most of the structure of the 21-parameter model. The 13-parameter model, shown in Figure 6.6, is also a subset of the 21-parameter model, with a simplified internal surface node and no ground heat transfer. The initial thought for this model is that for small footprint high-rise buildings the ground heat transfer may not be a significant contributor to the overall thermal load. Also a subset of the initial 21-parameter network, the 11-parameter model (Figure 6.7) contains the simplified internal surface network, as well as a simplified ground heat transfer network and lumped ceiling and exterior wall networks. The eight-parameter model, shown in Figure 6.8 further simplifies the 11-parameter model by neglecting ground heat transfer. This model contains a 3R2C network for exterior surfaces, a glazing resistance, and a simplified internal surface/mass network.

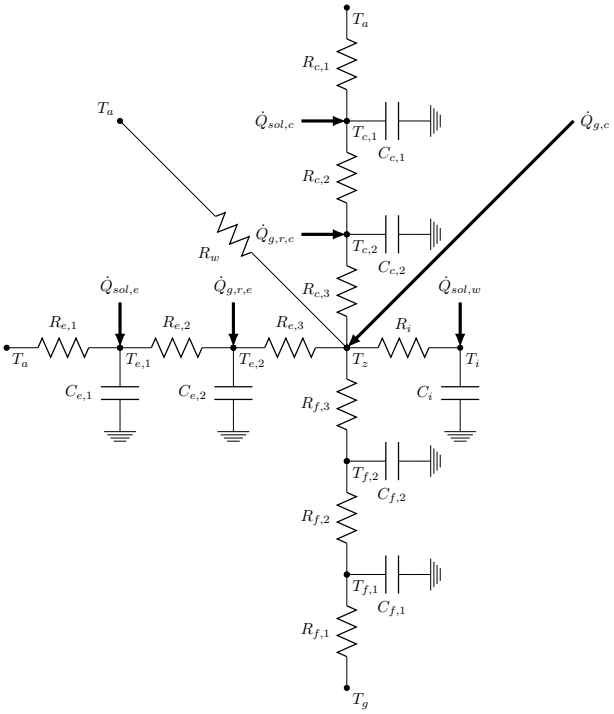

Figure 6.5: Eighteen-parameter thermal RC network.

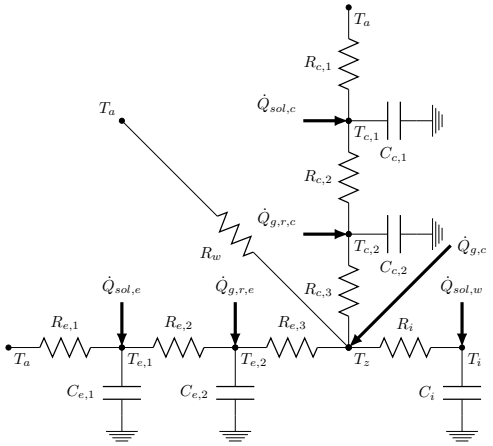

Figure 6.6: Thirteen-parameter thermal RC network.

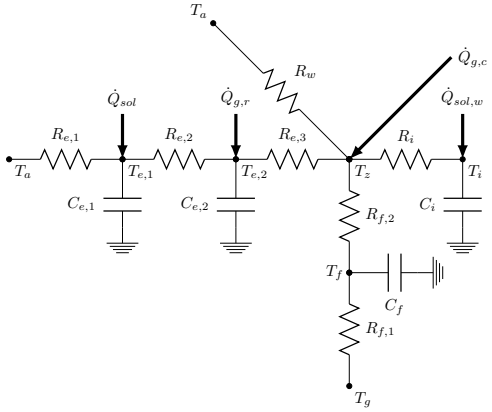

Figure 6.7: Eleven-parameter thermal RC network.

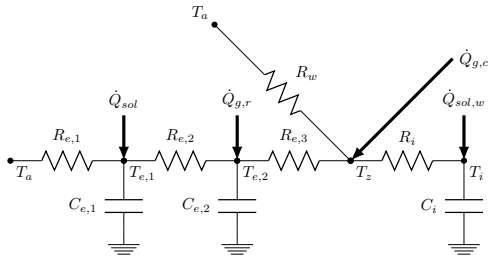

Figure 6.8: Eight-parameter thermal RC network.

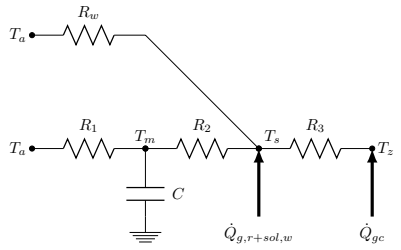

Figure 6.9: Five-parameter thermal RC network.

In this work, a five-parameter single-zone model is adopted, shown in Figure 6.9; its structure was adapted from the thermal RC network used in the CEN-ISO 13790 "Simple Hourly Method" load calculations [195]. Heat transfer and storage of opaque building shell materials are represented by R_1, R_2, and C. These elements link the ambient temperature node to a pseudo interior surface

temperature node T_s, accounting for potential heat storage of the mass materials. Glazing heat transfer is represented by a single resistance R_w connecting the ambient temperature node to the surface temperature node, because thermal storage of glazing is typically neglected. R_3 represents a lumped convection/radiation coefficient between the surface temperature node and the zone air temperature node T_z. The convective portions of internal gains (lighting, occupants, and equipment) are applied as a direct heat source to the zone temperature node, shown as $\dot{Q}_{g,c}$, and the radiant fraction along with glazing transmitted solar gains $\dot{Q}_{g,r+sol,w}$ are applied to the surface node.

It was decided to adopt the five-parameter thermal model after performing a model complexity analysis including all six thermal RC model structures previously presented. Each of the six RC networks (five-parameter, eight-parameter, 11-parameter, 13-parameter, 18-parameter, and 21-parameter) was trained using surrogate data from the five-zone U.S. Department of Energy Stand-alone Retail Reference EnergyPlus model. Table 6.2 summarizes the model performance in terms of root mean square error (RMSE) with respect to a validation dataset, and in terms of an objective generalized cross-validation score (GCV). GCV is defined in Equation 6.1 and essentially weights the mean-squared error based on model complexity [196].

Obviously, the 11-parameter model is superior to the five-parameter model in terms of RMSE and GCV. Visual inspection, however, proved that the model responses in terms of zone temperature and sensible zone load are virtually identical; therefore, the simpler five-parameter model were chosen to reduce sample size in the Monte Carlo analyses.

$$GCV = \frac{\sum_{i=1}^{N} (\dot{Q}_{rom,i} - \dot{Q}_{ep,i})^2}{N\left(1 - \frac{p}{N}\right)^2} \tag{6.1}$$

In Equation 6.1, N represents the total number of data points, $\dot{Q}_{rom,i}$ is the model predicted zone sensible load, $\dot{Q}_{ep,i}$ is the surrogate zone load, and p is the number of parameters in the model. The number of parameters p is equal to the number of RC parameters plus two, to account for the internal gain and zone capacitance multipliers that may also be used in model calibration.

For the retail building the 18-parameter model produced the lowest RMSE; however, the 21-parameter results were inadequate because they should have achieved at least the same score as the lower order model. The 11-parameter model produced the lowest GCV, which suggests that the additional improvement made by the 18-parameter model was not worth the additional complexity. Overall, the RMSE values for the retail building are all relatively low, further suggesting that all model forms performed well. Because satisfactory performance was observed from all models, the 5-parameter model was adopted to keep the problem dimensionality low.

Table 6.2: Model complexity results.

				Retail	
Model	p	N	k	RMSE	GCV
5p	7	504	128	6411	42.3×10^6
8p	10	504	1024	5338	29.7×10^6
11p	13	504	8192	3087	10.0×10^6
13p	15	504	32768	5234	29.1×10^6
18p	20	504	1048576	3076	10.3×10^6
21p	23	504	8388608	3192*	11.2×10^6

*Slightly suboptimal. Should have at least reached 3076 as the 18-parameter retail model.

A thermal RC network may be represented by a system of linear first-order differential equations with constant coefficients by performing an energy balance at each node with a storage element. This system can be represented in traditional state-space form as:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

For the five-parameter model adopted in the retail building modeling effort, state and input vectors are represented as:

$$\mathbf{x}^T = [T_m \;\; T_s]$$

$$\mathbf{u}^T = [T_z \;\; T_a \;\; \dot{Q}_{g,r+sol,w} \;\; \dot{Q}_{g,c}]$$

where T_m is the opaque building shell thermal temperature, T_s is the (pseudo) internal surface temperature, T_z is the zone temperature set point, T_a is the ambient external temperature, $\dot{Q}_{g,r+sol,w}$ is the sum of the radiative portion of internal gains and the solar radiation transmitted through glazing, and $\dot{Q}_{g,c}$ is the total convective internal gains.

The state space equations are then converted to the following heat transfer function form:

$$\dot{Q}_{sh,t} = \sum_{k=0}^{n} \mathbf{S}_k^T \mathbf{u}_{t-k\Delta\tau} - \sum_{k=1}^{m} e_k \dot{Q}_{sh,t-k\Delta\tau} \tag{6.2}$$

where $\mathbf{S}$ is a matrix containing input coefficients, e_k is a vector containing heat gain history coefficients, n is the number of past inputs in the calculation, and m is the number of past heat gain values in the calculation.

The transfer function method is an efficient calculation routine as it relates the sensible heat gains to the space ($\dot{Q}_{sh}$) at time t to the inputs ($\mathbf{u}_t$) of n and heat gains ($\dot{Q}_{sh,t}$) of m previous time steps. The input weighting coefficients ($\mathbf{S}_k^T$) and zone load coefficients (e_k) are the results of the state space to transfer function conversion process described by Seem [197].

Performing an energy balance on the zone air node in Equation 6.3 provides a basis for sensible zone load calculations where C_z is the zone air (or node) capacitance, T_z is the zone air temperature node, $\dot{Q}_{sh,t}$ is the zone-sensible heat gain, $\dot{Q}_{inf}$ represents infiltration heat gain, and $\dot{Q}_{zs,t}$ is the required sensible zone load. In effect, the RC network model describes the transient heat transfer through opaque and transparent envelope components as well as internal gains from occupants, lighting, and equipment. This network is used to compute the heat gains from these sources to the air node. The complete energy balance is provided in Equation 6.3, including infiltration, zone air mass, and HVAC heat addition and extraction rates.

$$C_z \frac{dT_z}{dt} = \dot{Q}_{sh,t} + \dot{Q}_{zs,t} + \dot{Q}_{inf} \tag{6.3}$$

If the differential in Equation 6.3 is approximated by:

$$\frac{dT_z}{dt} \approx \frac{T_{z,t} - T_{z,t-\Delta}}{\Delta\tau}$$

it can be rearranged to develop an algebraic "inverse" transfer function for computing zone temperature predictions from a known zone load shown in Equation 6.4.

$$\bar{T}_z = \frac{\sum_{l=2}^{r} \mathbf{S}_0(l)\mathbf{u}_t(l) + \sum_{j=1}^{n} \mathbf{S}_j \mathbf{u}_{t-j\Delta\tau} - \sum_{j=1}^{m} e_j \dot{Q}_{sh,t-j\Delta\tau} + 2\frac{C_z}{\Delta\tau} T_{z,t-\Delta\tau} + \dot{m}_{inf} C_p \mathbf{u}_t(2) + \dot{Q}_{zs,t}}{2\frac{C_z}{\Delta\tau} - \mathbf{S}_0(1) + \dot{m}_{inf} C_p} \tag{6.4}$$

where r is the number of inputs in input vector $\mathbf{u}$, here $r = 4$. An assumption of this formulation is that the heat gains are computed using the average value over the time step so the actual temperature at a given time step can be determined from:

$$T_{z,t} = 2\bar{T}_{z,t} - T_{z,t-\Delta\tau}$$

An ideal load calculation scheme for a dual set point with a dead band scenario can be described by using the previous equations according to the following procedure:

for t = simstart : simend do

Calculate $\dot{Q}_{sh,t}$ using Equation 6.2;

Calculate $\dot{Q}_{zs,t}$ to maintain $T_z = T_{cool,set}$ using Equation 6.3 (assume cooling first);

if $\dot{Q}_{zs,t} < 0$ **(heating required to maintain cooling set point) then**

Set $\dot{Q}_{zs,t} = 0$. Compute floating temperature using Equation 6.4;

if $T_z < T_{heat,set}$ **then**

Recompute $\dot{Q}_{zs,t}$ to maintain $T_z = T_{heat,set}$ using Equation 6.3

end

end

end

To compute zone humidity, the simulation also includes a moisture balance as described in Equation 6.5:

$$m_{air}\frac{dW_Z}{dt} = \dot{m}_{inf}(W_{OA} - W_Z) + \dot{m}_{SA}(W_{SA} - W_z) + \frac{q_{occ,lat}}{h_{fg}} \tag{6.5}$$

where m_{air} is the mass of air in the zone, $\dot{m}_{inf}$ is the mass flow rate of air from infiltration, $\dot{m}_{SA}$ is the supply air (SA) mass flow rate, $q_{occ,lat}$ is the occupant latent gain, and h_{fg} is the heat of

vaporization of water. W_z, W_{OA}, and W_{SA} are the humidity ratios of the zone, outdoor air, and SA, respectively.

6.3.1.3 Envelope Model Calibration

When using the inverse gray-box thermal modeling approach, it was necessary to determine the values of R and C parameters that bring the simple model into the closest agreement with the more detailed EnergyPlus model. Sum of squares minimization was used to identify model parameters that minimize the RMSE, defined by Equation 6.6, between the ROM predicted ($\dot{Q}_{rom}$) and the surrogate or measured ($\dot{Q}_{ep}$) sensible zone load.

$$J = \sqrt{\frac{\sum_{i=1}^{N} (\dot{Q}_{rom,i} - \dot{Q}_{ep,i})^2}{N}} \tag{6.6}$$

In this work, the two-stage optimization presented by Braun and Chaturvedi [188] was implemented that first performs a direct search over the parameter space to identify a starting point for local refinement. The direct search is performed on k uniformly random points located within the bounds of the parameter space. The local refinement, subject to the same parameter constraints, is performed via nonlinear least squares minimization implemented using the MATLAB optimizer `lsqnonlin` based on trust-region Newton methods. The implementation in this environment also allows local refinement to be performed around several good starting points from the direct search. For higher complexity models the local optimization can be sensitive to the initial starting point. Good results have been found when the 12 best direct search points are given to separately executed least squares algorithms to simultaneously explore several attractive regions. Table 6.3 presents the calibrated parameters for the five-parameter model used throughout this work. The zone air capacitance multiplier CM represents furnishing and other mass associated with the air node; the internal gain multiplier scales the assumed internal gains from lights and equipment. These were considered the nominal parameter values to which uncertainty was applied later in the work.

Table 6.3: Calibrated Five-Parameter Network RC Parameters.

Parameter	Value	Units
R_1	4.989	$(m^2K)/W$
R_2	0.164	$(m^2K)/W$
R_3	0.183	$(m^2K)/W$
R_w	3.000	$(m^2K)/W$
C	279.6	$kJ/(m^2K)$
CM	3.5	-
GM	0.788	-

6.3.1.4 HVAC System Modeling

For the standalone retail building a typical constant volume packaged rooftop unit (RTU) was modeled. Figure 6.10 provides an overview of the system configuration. The RTU model features a temperature- or enthalpy-based outdoor air economizer, constant-volume fan, single-speed direct expansion (DX) cooling coil, and gas heating coil. Component models were based on the quasi-steady-state physical formulations used by several mainstream whole-building simulation programs [198, 199]. Component models were programmed such that a full air loop can be simulated, allowing system air states to be included in a zone moisture balance for computing zone humidity levels.

Next, the fidelity of the new HVAC models were assessed against the EnergyPlus model. EnergyPlus outputs were used as inputs to the new HVAC models to compare with the HVAC system performance only. Comparing the EnergyPlus with the HVAC model implementations used in this work, Figures 6.11 and 6.12 show annual SA temperature and humidity ratio, respectively, for an annual simulation. In Figure 6.11, the top panel shows the SA temperature for occupied and unoccupied periods. To better visualize the information, weekly comparison plots are offered for a winter week and a summer week. Overall performance is very good during summer conditions. Mostly slight temperature deviations were noted during winter periods; however, early morning startup periods are visible where the ROM shows SA temperature values that are 10 K higher than the values found by EnergyPlus. After further review it was discovered that this is an artifact of

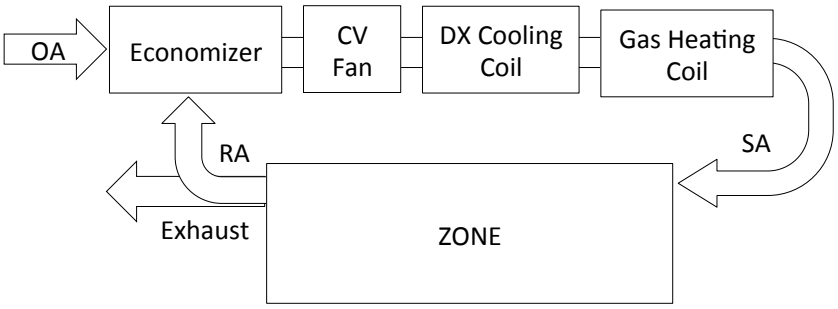

Figure 6.10: Packaged RTU.

using average hourly rather than subhourly EnergyPlus outputs as validation data, and the fact that the ROM was simulated at hourly time steps.

During unoccupied periods the fan and heating coil cycle ran in unison to meet the required heating loads. During this operating mode the RTU model reports the SA temperature as the air temperature leaving the heating coil, which is near 50°C when the coil is operating. In the case of the ROM, this temperature is reported for the entire hour even though the RTU does not run constantly for the hour. Because the EnergyPlus model was simulated at subhourly time steps, time intervals existed where they were not necessary for the heating coil and fan to run, and thus much lower supply temperatures were reported for some time steps. Thus, the hourly average SA temperature reported by EnergyPlus was 10 K lower. Had detailed (i.e., subhourly) EnergyPlus outputs been plotted for validation, several higher spikes near 50°C would have been observed along with lower values near 20°C during the hour.

Figures 6.13 and 6.14 highlight the calculated return air (RA) temperature and humidity ratio. Slight differences in the RA humidity can be observed in the results of the simplified zone moisture balance. As with SA temperatures, this is likely an artifact of using average hourly rather than subhourly EnergyPlus outputs as validation data. Simple first-order methods were used to implement the moisture balance and may also contribute to numerical differences between the two models. However, overall the model is a good approximation. Figures 6.15 and 6.16 show the predicted RTU energy consumption for an annual simulation.

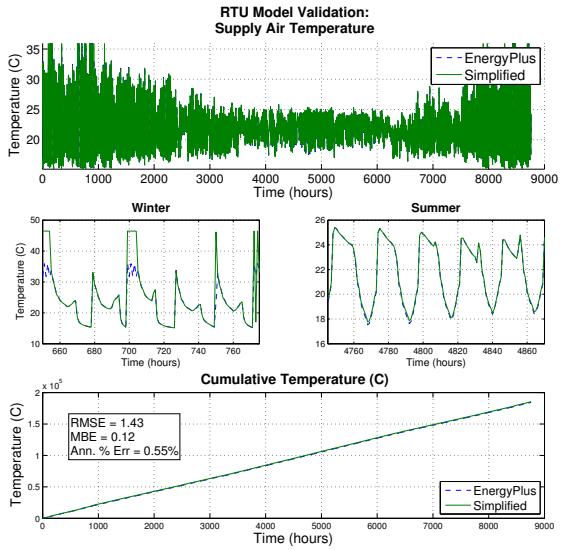

Figure 6.11: Validation of packaged RTU SA dry bulb temperature.

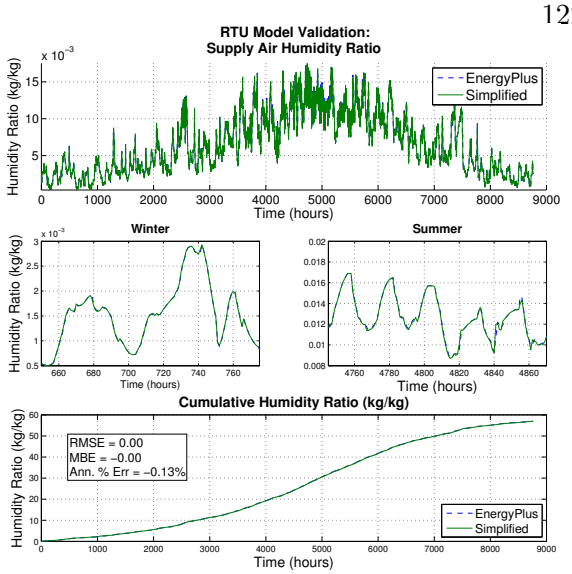

Figure 6.12: Validation of packaged RTU SA humidity ratio.

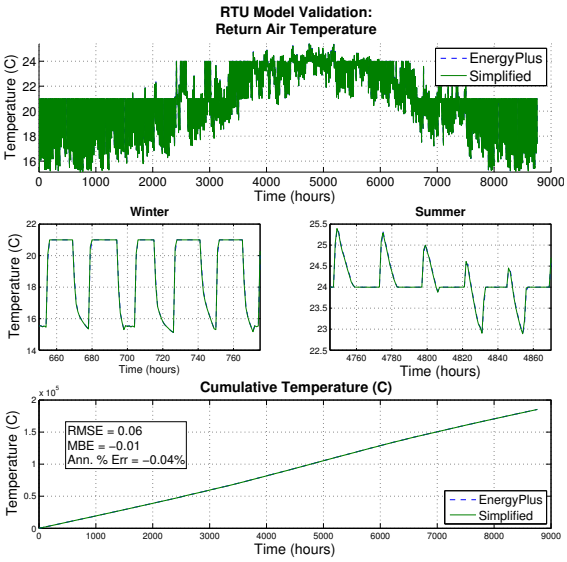

Figure 6.13: Validation of packaged RTU RA dry bulb temperature.

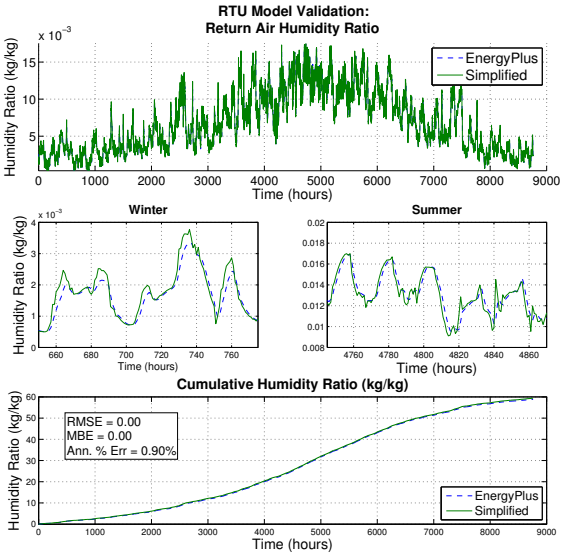

Figure 6.14: Validation of packaged RTU RA humidity ratio.

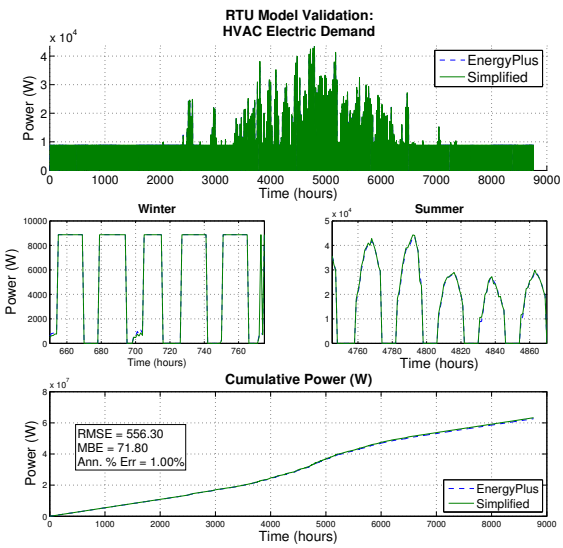

Figure 6.15: Validation of packaged RTU electricity demand.

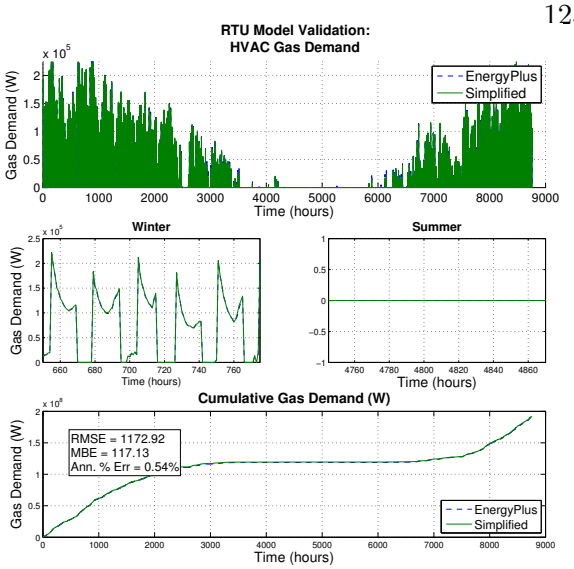

Figure 6.16: Validation of packaged RTU gas demand.

6.3.1.5 Overall Retail Building Model Validation

The following results provide a comparison of overall gray-box retail building model performance compared to its five-zone EnergyPlus counterpart. That is, the RTU HVAC models described and validated in Section 6.3.1.4 were coupled to the five-parameter thermal RC network that was developed in Section 6.3.1.2 and calibrated in Section 6.3.1.3, to evaluate the ROM in its entirety. To provide better insight into the model performance under various conditions, it was simulated using typical nighttime setup (NSU) operation during a mild week and a precooling heuristic for a hot week. (These are validation time periods; i.e., neither was included in model calibration.) Sensible zone load, temperature, and HVAC electricity consumption are in fairly good agreement for NSU and the precooling scenarios in Figure 6.17 through 6.22.

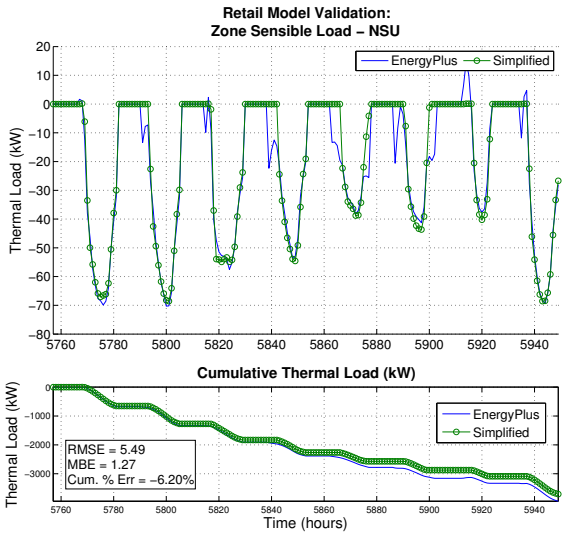

Figure 6.17: Retail sensible zone load comparison for NSU scenario.

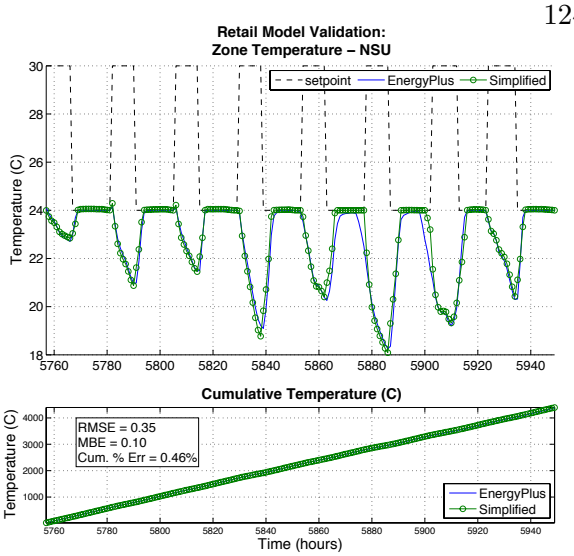

Figure 6.18: Retail zone mean air temperature comparison for NSU scenario.

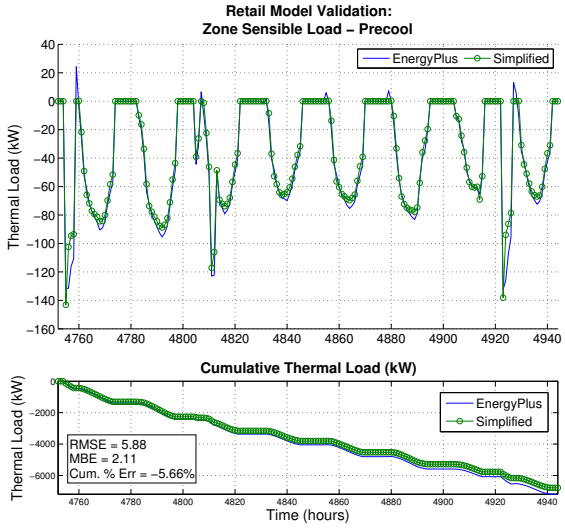

Figure 6.19: Retail sensible zone load comparison for precooling scenario.

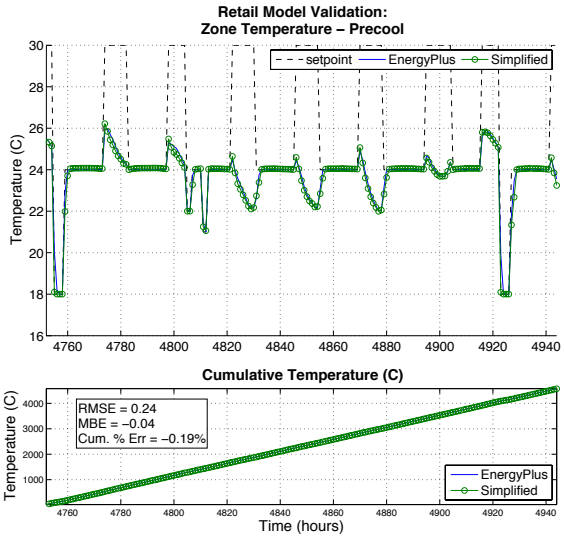

Figure 6.20: Retail zone mean air temperature comparison for precooling scenario.

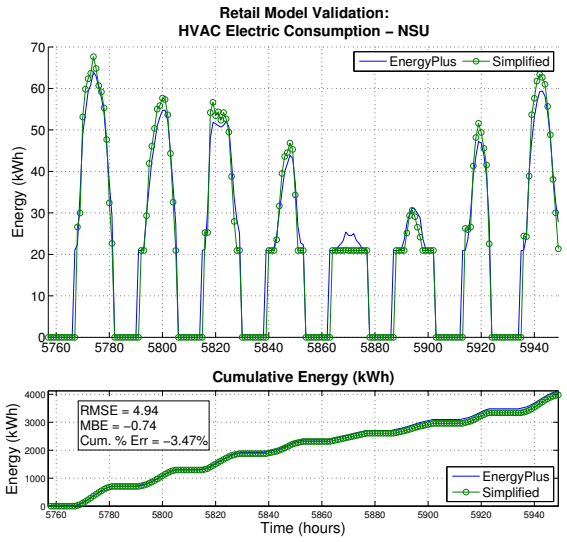

Figure 6.21: Retail HVAC electricity consumption comparison for NSU scenario.

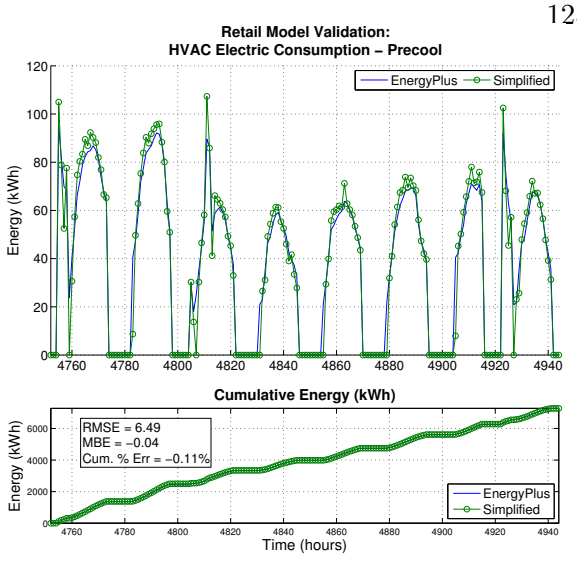

Figure 6.22: Retail HVAC electricity consumption comparison for precooling scenario.

6.3.2 Uncertainty Quantification

To illustrate the capabilities of the energy signal tool, the five-parameter envelope (single-zone) ROM was adopted; its parameters were identified from hourly surrogate training data derived from an EnergyPlus simulation of a five-zone retail building near the Chicago Midway Airport. This model has five parameters for the building shell; however, 20 parameters are required for the building, its use, and HVAC systems; each parameter is considered to be uncertain.

6.3.2.1 Model Input Parameter Uncertainty

Input parameter distributions are characterized in this work using Gaussian distributions; faults are modeled with triangular distributions, although any other probability distribution may be selected. In this work, uncertainties are known varieties that are correctly quantified by an energy analyst using input distributions of choice; faults are effects of unobserved uncertainties that affect the measured building performance but not the modeled predictions.

The chosen distributions represent the best available knowledge of each uncertain model parameter. Input parameters are distributed around a mean that equals the nominal parameter

value found from the parameter estimation process that has resulted in the validated ROM presented above. A standard deviation of 10% of the mean is adopted for the uncertain input parameters.

Adopting the five-parameter ROM, 11 parameters are associated with the building shell and use and an additional nine parameters are associated with the HVAC system. The seven nominal building envelope parameters are shown in Table 6.3, the four use parameters are shown in Table 6.4, and the nine nominal HVAC-related model parameters are shown in Table 6.5.

Table 6.4: Retail Building Use Parameters.

Parameter	Value	Units
Lighting power density	32.30	W/m^2
Equipment power density	5.23	W/m^2
Occupant density	7.1	m^2/per
Infiltration flow rate	3.9	m^3/s

Table 6.5: Retail Building HVAC Parameters.

Parameter	Value	Units
Supply fan efficiency	57	%
Maximum supply fan airflow	13.5	m^3/s
Supply fan pressure rise	883	Pa
DX coil rated capacity	319	kW
DX coil rated sensible heat ratio	70	%
DX coil rated coefficient of performance (COP)	3.2	-
DX coil rated air mass flow rate	16.0	kg/s
Gas heating coil-rated capacity	457	kW
Gas heating coil efficiency	80	%

To demonstrate the tool, five of the 20 input parameters are considered uncertain: lighting power density, equipment power density, occupant density, DX coil-rated COP, and gas heating coil efficiency. The remaining 15 input parameters are taken at their nominal values. In reality, most of these 20 parameters would be uncertain at different levels of uncertainty.

6.3.2.2 Operational and Equipment Faults

In spite of an energy analyst's best intentions and a belief that the model accurately reflects the building with all uncertainties, undetected faults may affect building performance. In this work, a fault differs from a parameter uncertainty in that it is present without the knowledge of the energy analyst or building operator; i.e., it is unobserved, yet still affects the measured conditional distributions of building energy end use.

An additional six model parameters were chosen to represent such a faulted state. Tehse are modeled by triangular distributions, which may be asymmetrical. The outdoor air fraction's nominal value is 24.5%, but may vary from 0% to 100%. The airside economizer high limit temperature is nominally 28°C, but can vary between 12 and 40°C. Essentially, at very low values of the high limit temperature the economizer is disabled; at high values the economizer operation is not overridden. Heating and cooling temperature set points may differ from their assumed values with a random offset from -1.5 to +1.5 K. Moreover, the internal gains schedules may be expanded by up to 3 hours or contracted by up to 3 hours from the nominal building operation schedule. Finally, the schedules for internal gains from lighting and equipment may be shifted by up to 4 hours in each direction; i.e., to earlier and later onsets. Because the environment uses hourly time steps, the schedule contraction/expansion and shift are rounded to the nearest full hour. For the fault-free scenario the six fault parameters are kept at their nominal values. For the faulted case, all six fault parameters are randomly perturbed concurrently; i.e., a very wide range of fault combinations is explored.

Table 6.6: Fault Ranges.

Parameter	Nominal	Minimum	Maximum	Units
Outdoor air fraction	24.5	0	100	%
Airside economizer high limit	28	12	40	°C
Cooling temperature set point	22.8	-1.5	+1.5	°C
Heating temperature set point	21.7	-1.5	+1.5	°C
Internal gains contraction/expansion	10	-3	+ 3	h
Internal gains shift	8am-6pm	-4	+ 4	h

Random sampling of 10,000 annual samples for each case is presented here. Because parameter combinations randomly drawn from the input distributions described above may lead to infeasible model configurations for which no solution can be computed, only valid model results are retained. The sample rejection was put in place to deal with parameter combinations that cause simulation errors such as non-convergence and subsequent crashes. Crashes are most often observed when randomly sampling a wide range of HVAC equipment-rated parameters, because the random sampling may not always produce physically consistent rated conditions. Thus, the wider the input parameter distributions, the more frequently invalid model results are generated. In the fault-free scenario 99.94% of all samples were valid (six of 10,000 failed); in the faulted scenario 99.88% of all samples were valid (12 of 10,000 failed).

The resultant conditional distribution of whole-building, HVAC, lighting (LTG), and plug electricity EQP consumption is shown in Figure 6.23; Table 6.7 shows pertinent statistics of the same in units of [MWh/a]. In the figure and table, whole-building electricity (WBE) please define and add to nomenclature is the sum of HVAC, LTG, and miscellaneous electric loads EQP. For fault-free and faulted cases, the table shows the mean, 10th, 50th (median), and 90th percentiles, along with the deterministic mean without consideration of any parameter uncertainty.

It is evident that the uncertainty associated with the five selected parameters only slightly affects the mean but strongly affects the variance in the whole-building and submetered end uses. The central 80% of the WBE consumption can be found within -7% and +8% from the deterministic mean in the fault-free case, and within -24% and +20% from the deterministic mean in the faulted case. The influence of faults is stronger in the HVAC and LTG end uses compared to the EQP end use. Given the dominance of HVAC and LTG end uses on whole-building consumption WBE, faults strongly affect WBE as well. Above all, faults widen the energy distributions.

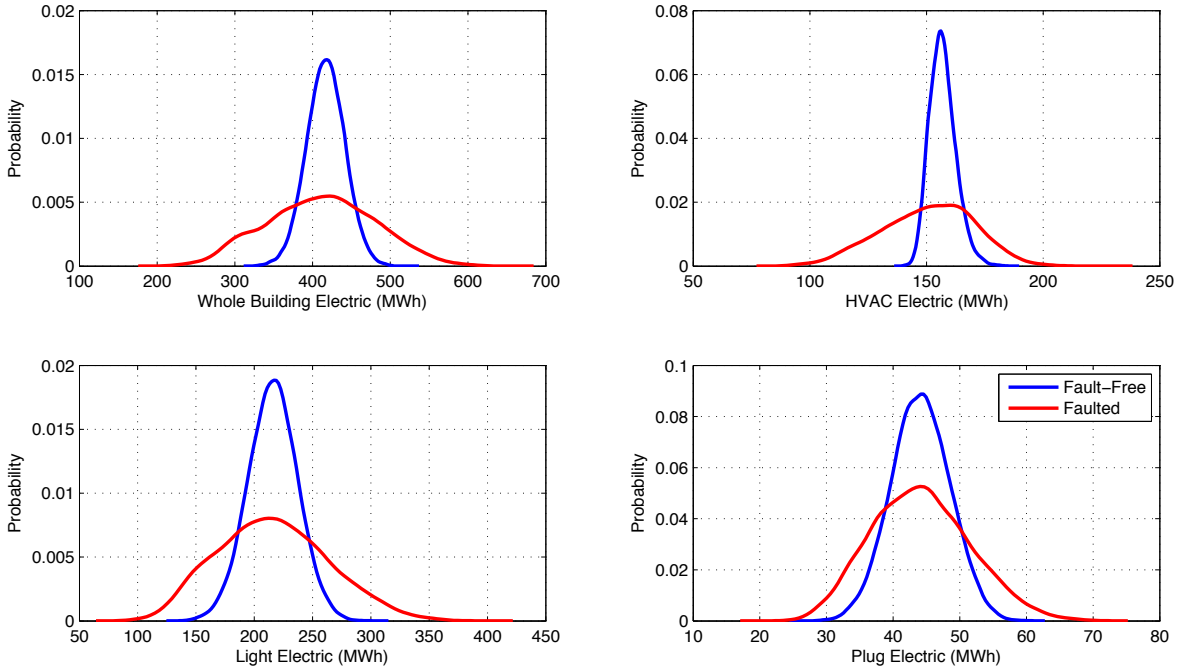

Figure 6.23: Distributions of whole-building, HVAC, lighting, and plug electricity consumption.

Table 6.7: Summary Statistics of Conditional Distributions of Whole-Building, HVAC, Lighting, and Plug Electricity Consumption in [%] in [MWh/a].

			Percentile		
End-Use	Case	Mean	10th	50th	90th
WBE	Deterministic	416.3			
	Fault-free	100%	93%	100%	108%
	Faulted	99%	76%	99%	120%
HVAC	deterministic	156.5			
	Fault-free	100%	96%	100%	105%
	Faulted	96%	79%	97%	112%
LTG	deterministic	215.8			
	Fault-free	100%	87%	100%	113%
	Faulted	100%	71%	99%	130%
EQP	deterministic	44.1			
	Fault-free	100%	87%	100%	113%
	Faulted	100%	78%	100%	123%

6.3.3 Decision Analysis

Comparing monitored data against a probable range of expected energy use is more insightful than comparing against a single number, because it allows a building owner to assess the urgency of corrective actions that need to be taken. If the measured energy use lies at the edge of the probable range of expected values, given all the uncertainties in the model inputs, the owner can be very confident that an issue requires attention. In this work, a decision-making tool was developed based on the probability distribution of model predictions to determine the expected utility of a range of available decisions, suggesting the one that maximizes the expected utility. The tool takes on the form of a modified traffic light with red, yellow, and green lights. The perspective that a red light is shown both for high levels of overconsumption and high levels of underconsumption

was adopted, because a building consuming significantly less energy than expected may indicate an operational problem as significant as a building consuming too much. A yellow signal is similarly used for cases of mild overconsumption and mild underconsumption. A green light is reserved for measured building energy consumption that is in line with model expectations.

As a first step, a distribution of the model-predicted energy consumption (called E_{mod}) was generated using Monte Carlo simulation as described in Section 6.3.2.

Second, the cases falling on the 5th, median, and 95th percentiles of the modeled energy distributions were selected to represent low, medium, and high estimates of actual measured energy consumption (called E_{meas}) for fault-free and faulted scenarios. The decision analysis compared the distribution of model-predicted energy consumption with E_{meas} to determine appropriate actions; thus, to illustrate the decision tool these three values of E_{meas} were taken from the sampled distributions. When physically implemented, the measured energy consumption would be determined directly from building metering data.

Third, boundaries were computed from the measured energy consumption to define meaningful ranges of low and high levels of deviation in energy consumption. Beginning with a low level of deviation, let us define $E_{0,low}$ such that it is X_{low} percent below E_{meas} and $E_{1,low}$ such that it is X_{low} percent above E_{meas}.

$$E_{0,low} = E_{meas}(1 - X_{low})$$

$$E_{1,low} = E_{meas}(1 + X_{low})$$

Similarly, for a high level of deviation, let's define $E_{0,high}$ such that it is X_{high} percent below E_{meas} and $E_{1,high}$ such that it is X_{high} percent above E_{meas}.

$$E_{0,high} = E_{meas}(1 - X_{high})$$

$$E_{1,high} = E_{meas}(1 + X_{high})$$

Of course, $X_{low} < X_{high}$ and arbitrarily defined X_{low} to be 5% and X_{high} to be 10%, i.e., a small deviation around the metered end use is ±5% and a large deviation ±10%. It would be easy

to adopt different values for X_{low} and X_{high} for each building energy end use depending on its temporal variability. Once a conditional probability distribution of expected energy consumptions is in hand, various kinds of statements can be made. Here, to report that the actual energy use is much higher, somewhat higher, similar, somewhat lower, or much lower than anticipated is desired. Further, to assign costs to making correct and incorrect statements, and report the statement that has lowest expected cost is desired.

Fourth, the empirical cumulative distribution of expected energy consumption was used to find the cumulative probabilities for the anticipated energy consumption to be below $E_{0,high}$ (called P_1), between $E_{0,high}$ and $E_{0,low}$ (called P_2), between $E_{0,low}$ and $E_{1,low}$ (called P_3), between $E_{1,low}$ and $E_{1,high}$, (called P_4) and above $E_{1,high}$ (called P_5). Together, these probabilities form the state probability vector $\vec{P} = (P_1, P_2, P_3, P_4, P_5)^T$ as shown in Figure 6.24.

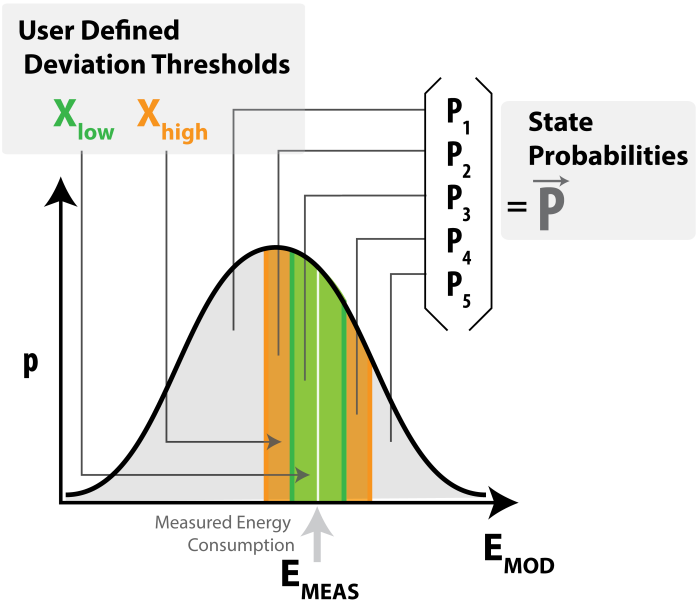

Figure 6.24: Relationship between deviation thresholds X_{low} and X_{high} and state probability vector $\vec{P}$.

Fifth, a cost function K is defined, where cost is a function of state and action with a finite number of states and a finite number of actions. Therefore, this cost function can be represented as

a matrix. Let us agree that there is one row per action and one column per state $K(i, j)$; i.e., cost of action i in state j. Let action $i = 1$ display the lower red light (RL) on the modified traffic signal, action $i = 2$ the lower yellow (YL) signal, $i = 3$ a green (G) signal, action $i = 4$ the upper yellow (YH) signal, and finally $i = 5$ be displaying the upper red (RH) light. Let $j = 1$ be the state that the model predicts a much lower (ML) energy consumption, $j = 2$ a somewhat lower (SL) energy consumption, $j = 3$ about the same (S), $j = 4$ a somewhat higher (SH) energy consumption, and $j = 5$ a much higher (MH) energy consumption than the actual building. Action vector $\vec{a}$ has thus five elements.

Table 6.8: Decision Analysis State and Action.

	State j: Anticipated Model Energy				
Action i:	Much Lower	Somewhat Lower	Similar	Somewhat Higher	Much higher
Lower red	K(RL,ML)	K(RL,SL)	K(RL,S)	K(RL,SH)	K(RL,MH)
Lower yellow	K(YL,ML)	K(YL,SL)	K(YL,S)	K(YL,SH)	K(YL,MH)
Central green	K(G,ML)	K(G,SL)	K(G,S)	K(G,SH)	K(G,MH)
Upper yellow	K(YH,ML)	K(YH,SL)	K(YH,S)	K(YH,SH)	K(YH,MH)
Upper red	K(RH,ML)	K(RH,SL)	K(RH,S)	K(RH,SH)	K(RH,MH)

An advantage of the presented decision analysis tool lies in its ability to individually set the costs of actions given certain states. Here, a reasonable but somewhat arbitrary assumptions on the values of the cost matrix elements: Showing an RL light when the model predicts ML energy consumption is assumed to have a cost of 4 (strong false negative, lost savings opportunity), when the model predicts an SL energy consumption a cost of 3 (false negative), when the model predicts a similar energy consumption a cost of 2 (weak false negative), when the model predicts an SH energy consumption a cost of 1 (weak false negative), and finally, when the model predicts MH energy consumption a cost of 0 (correct identification). Showing an RH light when the model predicts ML energy consumption is assumed to have a cost of 0 (correct identification), when the model predicts

an SL energy consumption a cost of 1 (weak false positive), when the model predicts a similar energy consumption a cost of 2 (weak false positive), when the model predicts an SH energy consumption a cost of 3 (false positive), and finally, when the model predicts MH energy consumption a cost of 4 (strong false positive, unnecessary alarm). Similar arguments can be made for the remaining signals of a lower yellow (YL), a green, and an upper yellow (YH) light, and yields the following cost matrix.

$$
\begin{aligned}
K &= \begin{bmatrix}
K(RL,ML) & K(RL,SL) & K(RL,S) & K(RL,SH) & K(RL,MH) \\
K(YL,ML) & K(YL,SL) & K(YL,S) & K(YL,SH) & K(YL,MH) \\
K(G,ML) & K(G,SL) & K(G,S) & K(G,SH) & K(G,MH) \\
K(YH,ML) & K(YH,SL) & K(YH,S) & K(YH,SH) & K(YH,MH) \\
K(RH,ML) & K(RH,SL) & K(RH,S) & K(RH,SH) & K(RH,MH)
\end{bmatrix} \\
&= \begin{bmatrix}
4 & 3 & 2 & 1 & 0 \\
3 & 2 & 1 & 0 & 1 \\
2 & 1 & 0 & 1 & 2 \\
1 & 0 & 1 & 2 & 3 \\
0 & 1 & 2 & 3 & 4
\end{bmatrix}
\end{aligned}
\tag{6.7}
$$

The expected cost vector for each action is found by multiplying the cost matrix K with the probability vector $\vec{P}$:

$$E(\vec{a}) = K \cdot \vec{P}$$

$$\begin{bmatrix} E(RL) \\ E(YL) \\ E(G) \\ E(YH) \\ E(RH) \end{bmatrix} = \begin{bmatrix} K(RL,ML) & K(RL,SL) & K(RL,S) & K(RL,SH) & K(RL,MH) \\ K(YL,ML) & K(YL,SL) & K(YL,S) & K(YL,SH) & K(YL,MH) \\ K(G,ML) & K(G,SL) & K(G,S) & K(G,SH) & K(G,MH) \\ K(YH,ML) & K(YH,SL) & K(YH,S) & K(YH,SH) & K(YH,MH) \\ K(RH,ML) & K(RH,SL) & K(RH,S) & K(RH,SH) & K(RH,MH) \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix}$$

$$= \begin{bmatrix} K(RL,ML)P_1 + K(RL,SL)P_2 + K(RL,S)P_3 + K(RL,SH)P_4 + K(RL,MH)P_5 \\ K(YL,ML)P_1 + K(YL,SL)P_2 + K(YL,S)P_3 + K(YL,SH)P_4 + K(YL,MH)P_5 \\ K(G,ML)P_1 + K(G,SL)P_2 + K(G,S)P_3 + K(G,SH)P_4 + K(G,MH)P_5 \\ K(YH,ML)P_1 + K(YH,SL)P_2 + K(YH,S)P_3 + K(YH,SH)P_4 + K(YH,MH)P_5 \\ K(RH,ML)P_1 + K(RH,SL)P_2 + K(RH,S)P_3 + K(RH,SH)P_4 + K(RH,MH)P_5 \end{bmatrix} \tag{6.8}$$

As suggested by utility theory, the last step is to select the best action a_{opt} in the face of uncertainty; i.e., activate that light, which minimizes the expected cost.

$$a_{opt} = \underset{i}{\operatorname{argmin}}\, E(\vec{a})$$

6.3.4 Bayesian Updating

Thus far, the distributions of the uncertain model parameters were assumed to be Gaussian with a standard deviation equal to a fixed fraction of the parameter mean, here 10%. The choice of these distributions was made somewhat arbitrarily, before any operational data was available from the actual building performance and termed before data because of its consideration prior to observation. It is believed that building performance measurements collected over an extended time period (i.e., after data) can be used to infer improved input parameter distributions by applying probability theory in general and Bayes' theorem in particular. See Jaynes [200] for full development of Bayesian probability. A brief explanation is provided below for information pertinent to this research.

High-dimensional integrals associated with problems in computational physics lead to the development of Markov Chain Monte Carlo algorithms, which can efficiently sample from probability distributions by exploiting the Markov property. This has led to the explosion of Bayesian

techniques, with the Metropolis Algorithm as the breakthrough approach (named as one of the top 10 algorithms of the 20th century) [201]. A probabilistic perspective not only provides insight into the relationship between sets of model parameters, revealing tradeoffs and compensating interactions, but also lends itself to a continuous model uncertainty quantification and tuning where the posterior distribution of an initial parameter estimate can be used as the prior for a subsequent parameter estimation update once new building performance data have been collected.

Dodier used Bayesian (belief) networks for whole-building energy diagnostics [36, 35]. Lauret et al. [202] demonstrated improvements over traditional parameter estimation methods by applying Bayes' theorem to determine better estimates of convection coefficients for a radiant barrier roof system model. More recently, Booth [203] used London housing stock models for hierarchical modeling, with considerations of internal heating set points, fraction of space heating, air leakage, heating system COP, window U-value, and window-to-wall ratio.

In this work, the Bayesian inference of uncertain model parameters relies on the extension of a previously developed technique [16]. It has benefits over traditional methods because prior knowledge of the system can be directly incorporated into the estimation task and methods for addressing sensor noise are inherent to the Bayesian approach. The inference can essentially be thought of as fitting a joint probability distribution to a measured dataset. Specifically, conditional probabilities are related through the product rule to derive Bayes' theorem and allow consideration of "before data" and "after data" states of knowledge. The prior probability distribution is updated with any measured data to form the posterior probability distribution, which represents the state of knowledge in any inference task. A periodic process where model input parameter distributions are updated daily, weekly, monthly, or for a similar period of interest is proposed. As new measurements become available, data effectively shape the distribution of expected building energy use according to the information gleaned from a combination of prior knowledge and sensor data; uncertainty is still present but should decrease with additional data and understanding of the relationships among variables.

The probability of parameter set Θ given measured data D and knowledge of the system

K can be written as posterior probability $p(\Theta|DK)$. Bayes' theorem then allows the conditional probability $p(\Theta|DK)$ to be computed from $p(\Theta|K)$, $p(D|\Theta K)$, and $p(D|K)$ as shown in Equation 6.9,

$$p(\Theta|DK) = p(\Theta|K)\frac{p(D|\Theta K)}{p(D|K)} \tag{6.9}$$

where $p(\Theta|K)$ represents prior knowledge about parameter values, $p(D|\Theta K)$ represents the likelihood of observing the measured dataset D given a particular parameter set Θ and knowledge of the system K, and $p(D|K)$ is the probability of observing the dataset. Ignoring the reference to system knowledge K, the relation can be written in alternate form where the numerator remains the product of likelihood and prior, and the denominator is a normalization factor so that posterior probabilities sum to unity.

$$p(\Theta|D) = \frac{p(\Theta)p(D|\Theta)}{\sum_i p(\Theta_i)p(D|\Theta_i)} \tag{6.10}$$

Assuming random Gaussian noise about a measured datum D_i, the likelihood of an observation can be determined from its location within the normal distribution with standard deviation σ_ϵ this is not in nomenclature, centered at μ equal to the measured datum,

$$p(D_i|\Theta) = \frac{1}{\sigma_\epsilon\sqrt{2\pi}}\exp\left(\frac{-(D_i - M_i)^2}{2\sigma_\epsilon^2}\right), \tag{6.11}$$

where M_i is the model output given the parameter set Θ.

Further, assuming independent errors, the likelihood is that the entire dataset is simply the product of the likelihoods of all individual points. The assumption is likely valid for common HVAC sensors (e.g., temperature probes), but correlated errors could be handled with a slightly different formulation that is indicative of a fault model. Measurement errors are often correlated because of, for example, hysteresis error or linearity error. If such correlated errors are of concern, a Bayesian (or other probabilistic) method may be used that can accommodate correlated measurements. This work nonetheless assumes uncorrelated energy consumption measurements; thus, it ignores autocorrelation of errors, which is estimated to be small. For the dynamics and time range

considered in this problem, model structure is considered more important, with respect to data fit, than noise correlation. From this model assumption, the easily computable likelihood function given by Equation 6.12.

The likelihood function is maximized when the exponential term is minimized, which occurs as the modeled data approach the measured (or surrogate) data. When uniform priors are used with Equation 6.12 in a Bayesian calibration context, the most likely parameters are equivalent to those that would be found using a least squares approach, because the exponential term in Equation 6.12 is essentially the sum of squared errors [152].

$$p(D|\Theta) = \frac{1}{\left(\sigma_\epsilon\sqrt{2\pi}\right)^n} \exp\left(\frac{-1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2\right) \tag{6.12}$$

Evaluating Equation 6.12 directly can pose numerical issues, because a small range of σ_ϵ values results in a large range of likelihoods. Double precision computing environments are typically capable of evaluating floating point numbers on the order of 10^{-308} to 10^{308}. This means that when using 3 weeks of hourly data (i.e., $n = 504$), σ_ϵ must approximately be in the range of $[0.1, 1.5]$. Values outside this range will cause the likelihood (and consequently the posterior) to evaluate to "Inf," "NaN," or "0," regardless of the time series fit. These numerical issues can be alleviated by computing the natural logarithm of the posterior rather than the posterior directly [202, 204].

To compute the natural log of the posterior, first, the log of both sides of Equation 6.10 is taken.

$$\ln\left(p(\Theta|D)\right) = \ln\left(\frac{p(\Theta)p(D|\Theta)}{\sum_i p(\Theta_i)p(D|\Theta_i)}\right) \tag{6.13}$$

The right-hand side of Equation 6.13 can be separated using logarithm product and quotient rules.

$$\ln\left(p(\Theta|D)\right) = \ln\left(p(\Theta)\right) + \ln\left(p(D|\Theta)\right) - \ln\left(\sum_i p(\Theta_i)p(D|\Theta_i)\right) \tag{6.14}$$

The log-likelihood term of Equation 6.14,

$$\ln\left(p(D|\Theta)\right) = \ln\left(\frac{1}{\left(\sigma_\epsilon\sqrt{2\pi}\right)^n}\exp\left(\frac{-1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2\right)\right) \tag{6.15}$$

can be further simplified by applying product and quotient rules as shown in Equations 6.16 and 6.17, respectively.

$$\ln\left(p(D|\Theta)\right) = \ln\left(\frac{1}{\left(\sigma_\epsilon\sqrt{2\pi}\right)^n}\right) + \ln\left(\exp\left(\frac{-1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2\right)\right) \tag{6.16}$$

$$\ln\left(p(D|\Theta)\right) = \ln\left(1\right) - \ln\left(\left(\sigma_\epsilon\sqrt{2\pi}\right)^n\right) + \ln\left(\exp\left(\frac{-1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2\right)\right) \tag{6.17}$$

With $\ln(1) = 0$, and the power rule can be applied to the middle term of the right-hand side. The last term of the right-hand side simplifies, because of to logarithmic identity, to produce Equation 6.18.

$$\ln\left(p(D|\Theta)\right) = -n\ln\left(\sigma_\epsilon\sqrt{2\pi}\right) + \frac{-1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2 \tag{6.18}$$

Recombining the simplified log-likelihood of Equation 6.18 with the log-posterior equation of Equation 6.14 yields:

$$\begin{aligned}\ln\left(p(\Theta|D)\right) = \ln\left(p(\Theta)\right) - n\ln\left(\sigma_\epsilon\sqrt{2\pi}\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2 \\ - \ln\left(\sum_i p(\Theta_i)p(D|\Theta_i)\right)\end{aligned} \tag{6.19}$$

The last term of the right-hand side of Equation 6.19 is ultimately a constant number subtracted from each individual $\ln(p(\Theta_i)p(D|\Theta_i))$ value. Because the value of this constant term does not impact the shape or relative information of the posterior, it could be thought of as an arbitrary constant C.

$$\ln\left(p(\Theta|D)\right) = \ln\left(p(\Theta)\right) - n\ln\left(\sigma_\epsilon\sqrt{2\pi}\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2 + C \tag{6.20}$$

The constant term can be moved to the left-hand side of the equation, producing Equation 6.21.

$$\ln\left(p(\Theta|D)\right) - C = \ln\left(p(\Theta)\right) - n\ln\left(\sigma_\epsilon\sqrt{2\pi}\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2 \tag{6.21}$$

Because the objective is to avoid numerical underflow or overflow, prescribing

$$C = \max\left(\ln\left(p(\Theta)\right) - n\ln\left(\sigma_\epsilon\sqrt{2\pi}\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(D_i - M_i\right)^2\right) \tag{6.22}$$

shifts all points so that the maximum is 0. A maximum value of 0 in the ln space ensures that all values will be mapped to the interval $[0, 1]$ when taking the exponential. After taking exponentials, the values can be scaled by a constant so that probabilities sum to unity.

The σ_ϵ value is a noise term with physical interpretation. Here there are three energy signals of interest: HVAC, lighting, and equipment–all electrical terms with the error associated with minor fluctuations not incorporated in the physical model. As previously stated, prescribing an appropriate σ_ϵ is necessary to prevent underflow or overflow, which causes the inference task to crash from numerical issues. The most appropriate σ_ϵ value can be found by maximum a posteriori (MAP) estimation. MAP is used to obtain a point estimate of σ_ϵ by placing a prior distribution over σ_ϵ and finding the maximum posterior mode according to the empirical data. Because the HVAC, lighting, and equipment energy signals are considered independent in this study, the MAP estimate of $\sigma_{\epsilon,\ HVAC}$, $\sigma_{\epsilon,\ LTG}$, and $\sigma_{\epsilon,\ EQP}$ were performed separately with a uniform prior on each σ_ϵ set between 0% and 15% of the magnitude of the full, individual signals.

With the appropriate and optimal σ_ϵ set for each signal, it was then possible to calculate the probability of observing various energy signals as a function of the uncertainty parameters. Because the signals are considered independent, the joint probability of the building state is a product of the individual probabilities:

$$p = \prod_i p_i = p_{HVAC}p_{LTG}p_{EQP}. \tag{6.23}$$

With the joint posterior probability distribution available from the equation above, it was then possible to sample from the posterior directly or marginalize over all parameters not of interest

and form a new prior. That is, the updating process is *prior* $\rightarrow$ *posterior* $\rightarrow$ *prior*. The optimal updating period is a function of the building energy dynamics and available data.

6.4 Results

Results are presented that (1) exemplify the decision support aspect of the tool; and (2) illustrate the updating of the uncertain model parameters based on measured data.

6.4.1 Decision Support Case Studies

The decision analysis results presented here are separated into fault-free and faulted scenarios. In the fault-free scenario, the high, medium, and low consumption values (surrogates of measured building energy consumption) are drawn from the conditional energy consumption distributions without faults and the decision analysis is based on the same distribution. In contrast, in the faulted scenario, the high, medium, and low consumption values are drawn from the conditional energy consumption distributions **including faults** and the decision analysis is based on the distribution **excluding faults**.

The results are shown as a matrix of figures. The first row shows the whole-building electricity WBE consumption results for the last year, then last month, then last week, followed by the last day. The second row shows the HVAC energy consumption results for the time periods, the third row shows the lighting LTG results, and lastly, the fourth row shows the EQP results. Each of the 16 figure panels reveals a box plot[2] Larger and smaller values, respectively, are shown as outliers of the expected energy consumption value for the time period of interest, a diamond marker superimposed on the box plot to indicate the surrogate actual consumption value E_{meas}, and on the left margin the energy signal tool with the signal chosen for the resultant cumulative probabilities $\vec{P}$, cost matrix C, and deviation thresholds X_{low} and X_{high}. The central green light separates cases of mild (YH) and strong (RH) overconsumption above the green light from the cases of of mild (YL) and

[2] Box plots shown adopt the common notation that the box occupies the interquartile range (IQR) from the lower (25th percentile) to the upper (75th percentile). The whiskers extend to the minimum and maximum values if these are less than 1.5 times the IQR below the lower or 1.5 times the IQR above the upper quartile.

strong (RL) underconsumption below the green light. The particular traffic signal-inspired design is one of many possible designs chosen for illustration here; thus, many other valid designs can be conceived. Moreover, the planned field implementation of this energy signal tool would likely not show the box plots but only the signals. Finally, if only whole-building energy WBE measurements are available, the tool would reveal only the top row of WBE versus the four time periods. In contrast, when submetering of HVAC, LTG, and EQP is available, the lower three rows would be shown and the top WBE row omitted, because it would not offer additional insight.

6.4.1.1 Fault-Free Scenario

High Energy Consumption Case Beginning on August 30 with the high energy consumption case at the 95th percentile, a mild overconsumption is shown for all time scales for WBE and HVAC; LTG and EQP show strong overconsumption for all time scales.

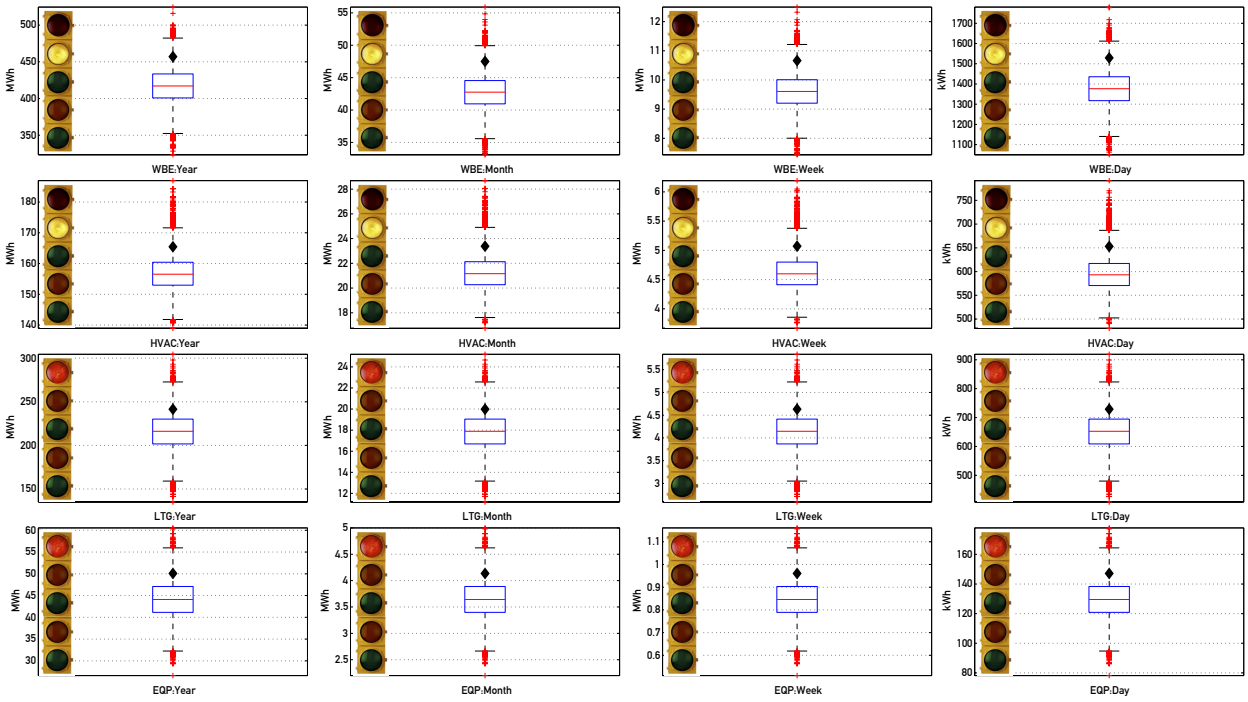

Figure 6.25: Fault-free high consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

Medium Energy Consumption Case The medium consumption cases for August and February show consistency between WBE, HVAC, and LTG; green lights are shown for all time scales and the EQP consumption is low. Because the EQP contribution to the total is small, the WBE signal is not swayed to show a low yellow light.

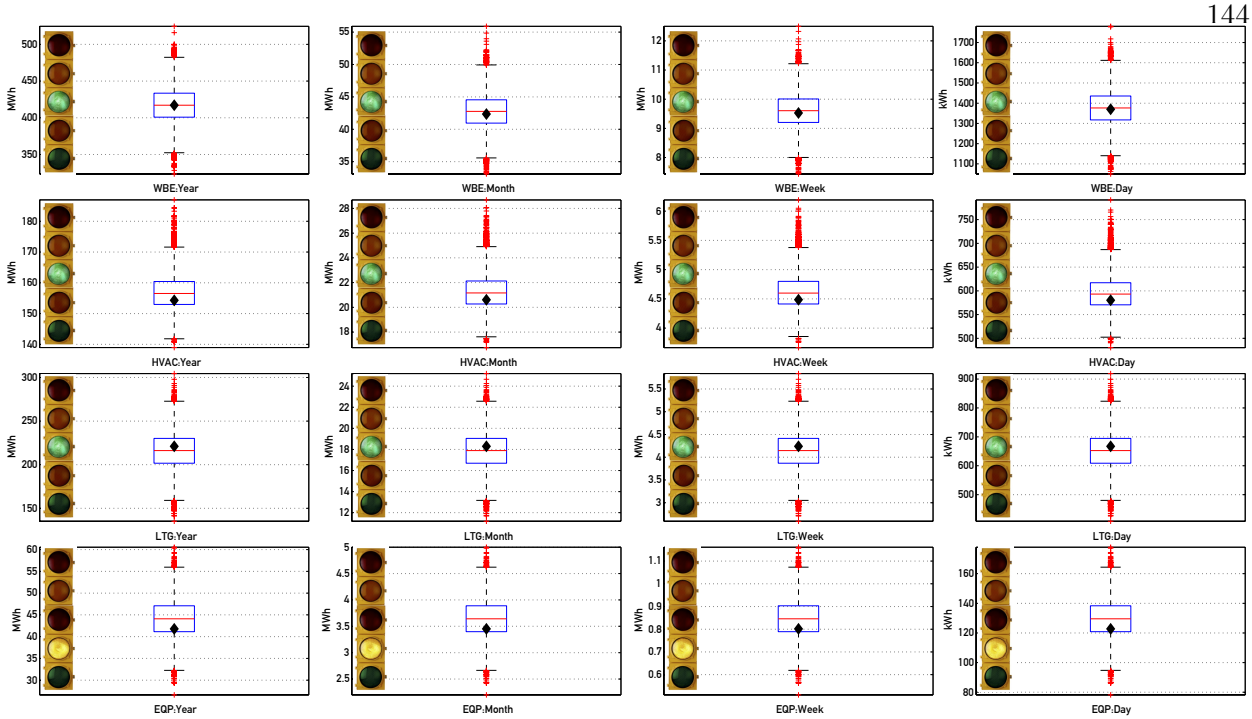

Figure 6.26: Fault-free medium consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

Low Energy Consumption Case In the low consumption case, it is interesting to note that the WBE signal shows a strong underconsumption (low red), driven by the corresponding LTG signal, even though HVAC consumption is similar to the model expectation (green) and EQP is only a mild underconsumption. This case reveals the importance of submetering: Without it, it would not have been possible to isolate that LTG causes the warning, HVAC could be ignored, and EQP could be given less consideration.

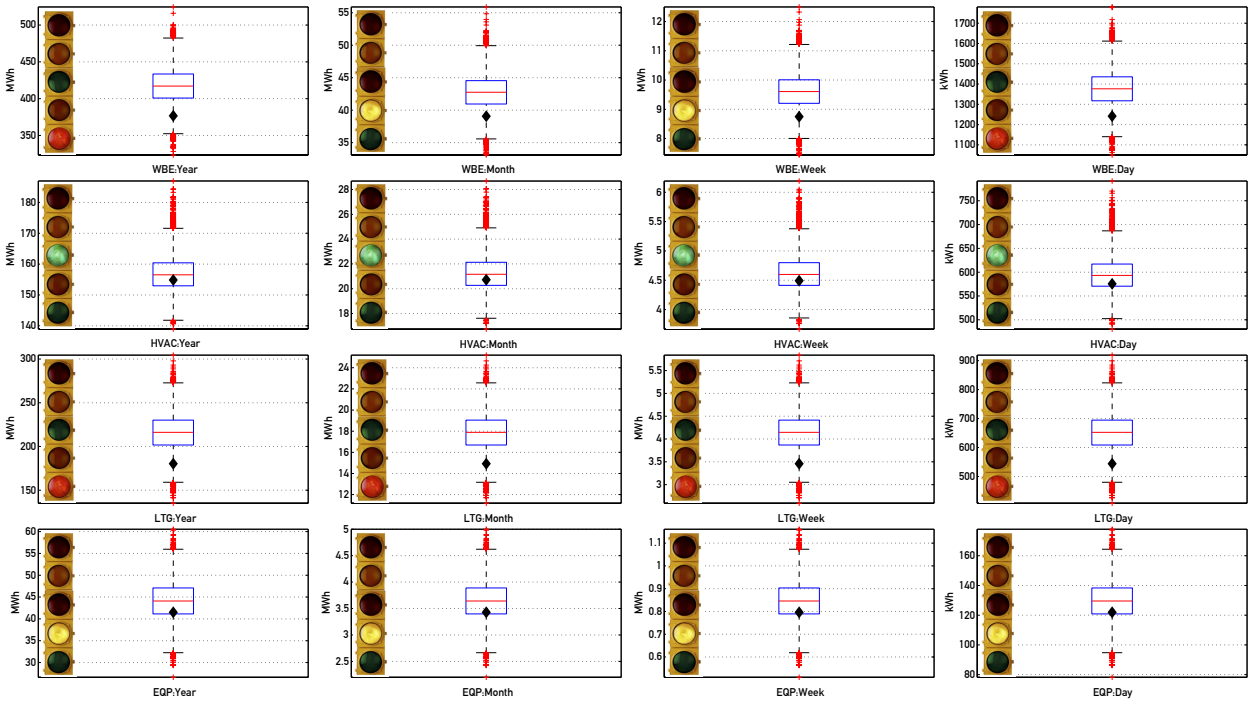

Figure 6.27: Fault-free low consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

6.4.1.2 Faulted Scenarios

In the faulted scenarios, the three cases are drawn from much wider distributions, as shown in Figure 6.23.

High Energy Consumption Case The high consumption case for August 30 shows consistency between WBE, HVAC, and LTG. RH lights are shown for all time scales and the EQP consumption is close to model expectation. However, because the EQP contribution to the total is small, the WBE signal is not swayed to show a low green light.

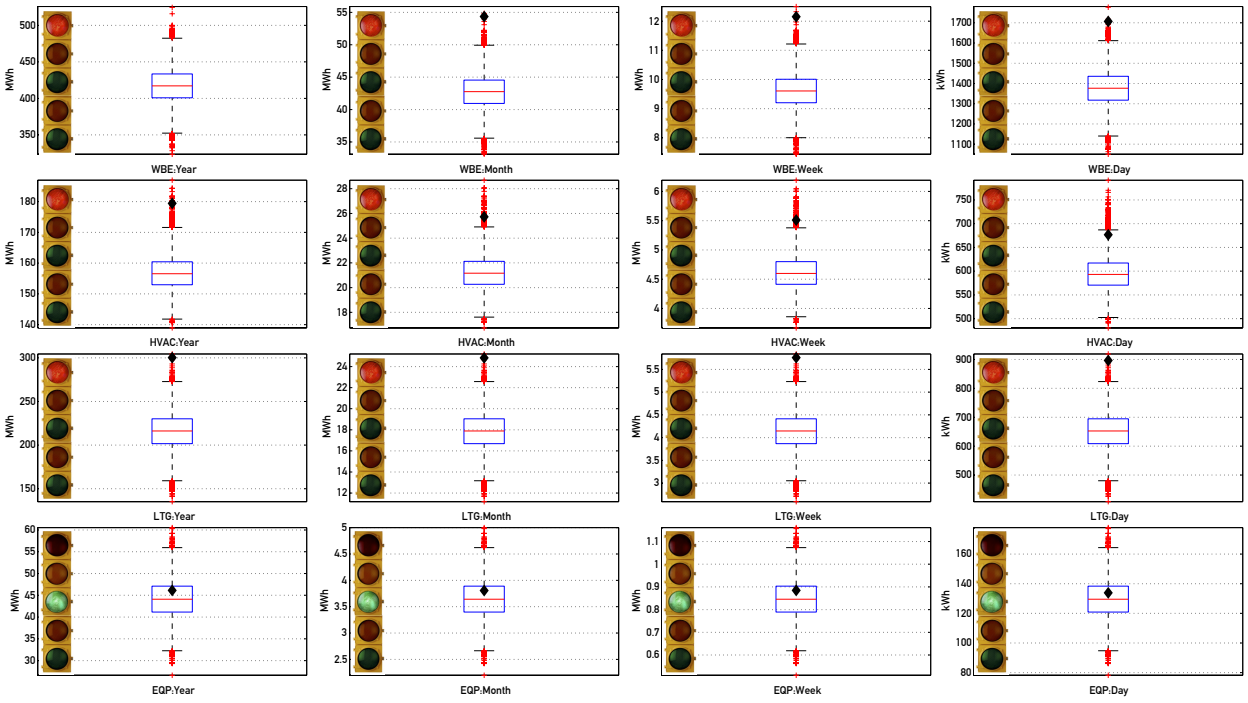

Figure 6.28: Faulted high consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

Medium Energy Consumption Case On an annual basis, the medium consumption case leads to the expected green lights. On time scales of months and shorter, we can observe HVAC mild and strong underconsumption.

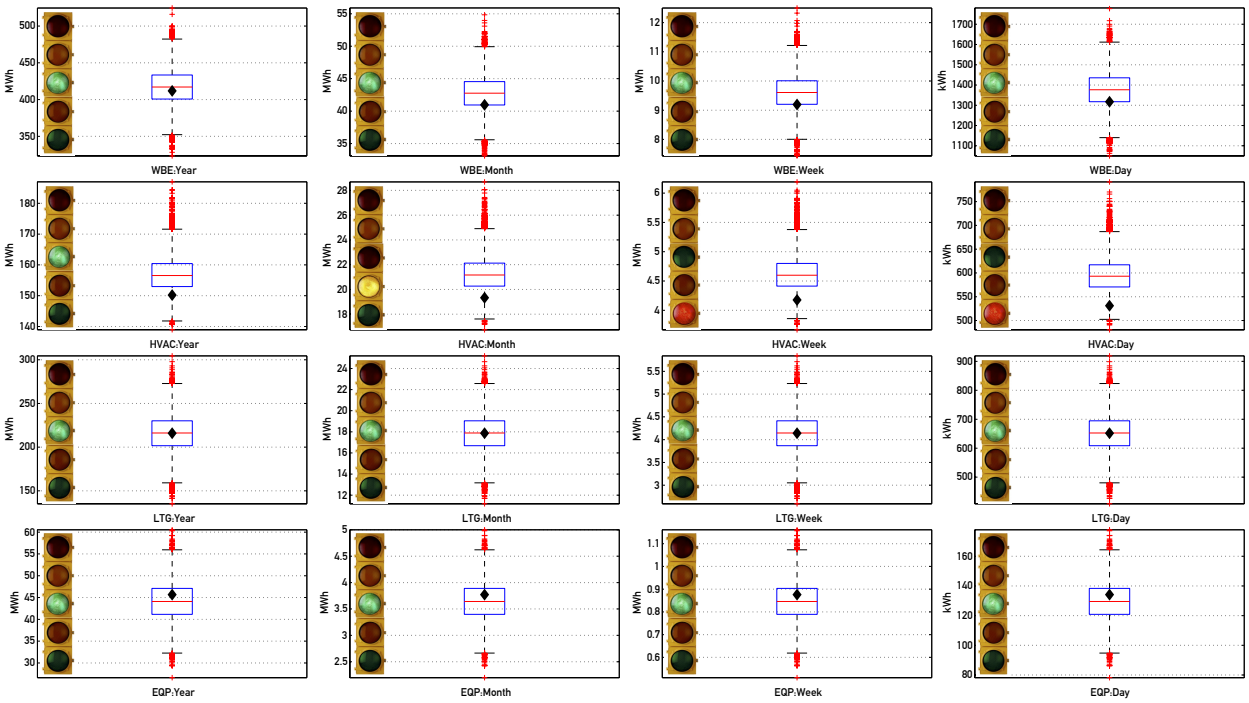

Figure 6.29: Faulted medium consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

Low Energy Consumption Case The low consumption case at the 5th percentile of the faulted distribution shows a consistent RL light for all end uses and time scales, independent of season.

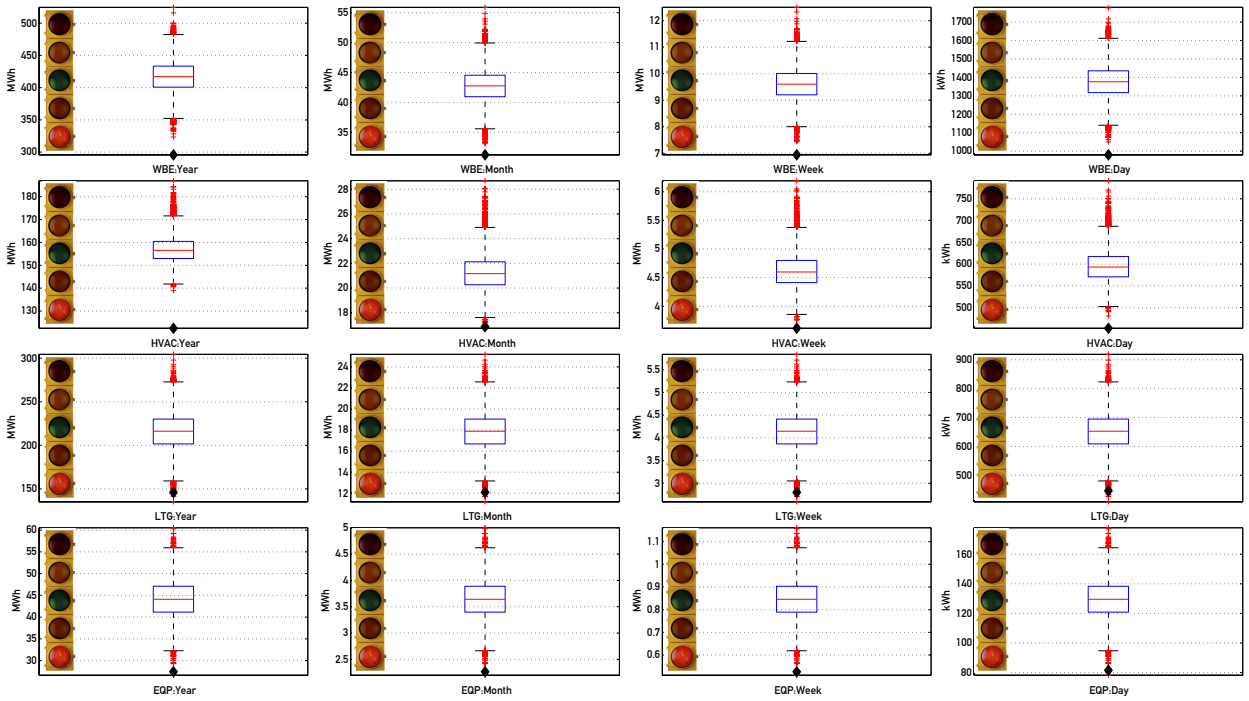

Figure 6.30: Faulted low consumption case beginning August 30 (measured consumption data are indicated by diamonds in each figure).

Finally, a sample illustration of how such a decision analysis tool would be deployed in the management of distributed commercial buildings is presented. Imagine, a building operator is responsible for the energy-efficient operation of four buildings in a city. Building 1 appears to be healthy; all end uses show green signals for all four time scales; building 2 suffers from overconsumption problems in the HVAC system that manifest themselves in the last day, week, and month; building 3 exhibits underconsumption in LTG, especially during the last week, which could speak to failed light sources or delamping measures not yet accounted for in the model; finally, building 4 suffers from multiple symptoms: overconsumption in HVAC and EQP, and underconsumption in LTG.

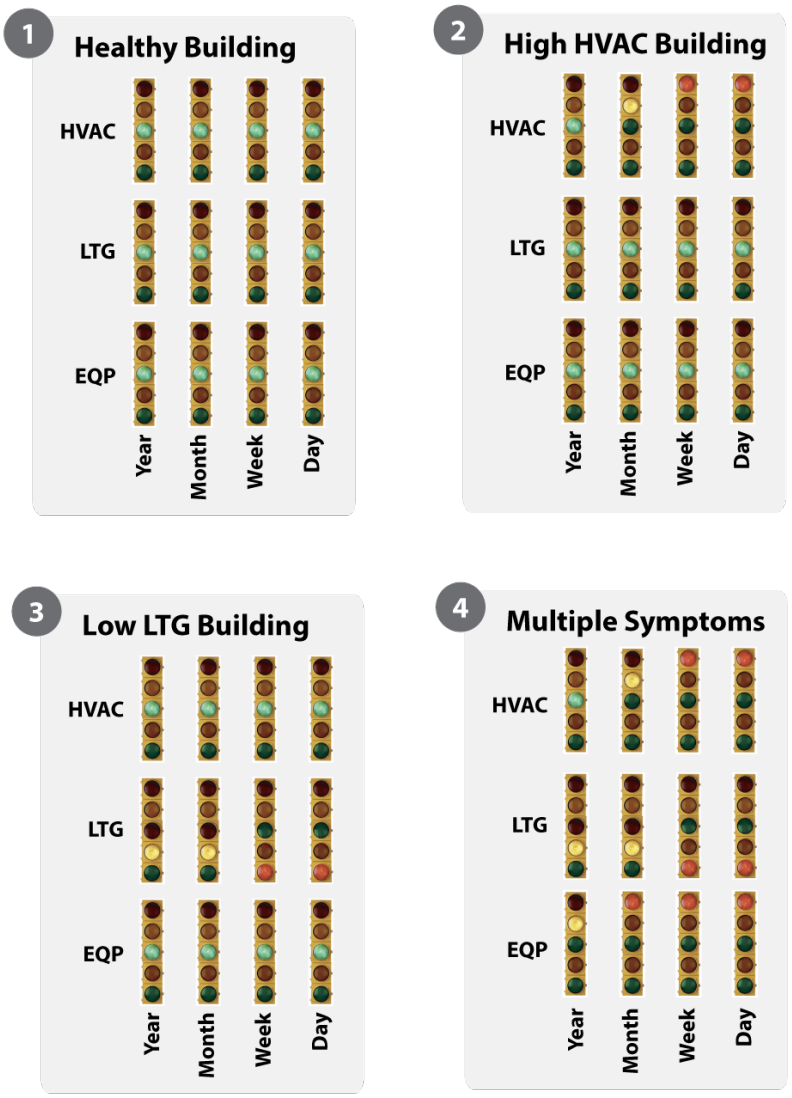

Figure 6.31: Example of four buildings managed by a building operator.

6.4.2 Bayesian Parameter Updating Case Studies

Results related to the probabilistic inference and updating of the uncertain model parameters are presented for September 21, at which point measurements over the past 30 days are used to update the five input parameter distributions that are deemed uncertain. The high consumption case representing the 95th percentile of the fault-free conditional distributions is used to compute the

likelihood functions for HVAC, LTG, and EQP for each of the nearly 10,000 parameter combinations. Based on the simplifying assumptions articulated above, the marginal posterior distribution for each selected uncertain parameter can be found. Please note that white Gaussian noise with a signal-to-noise ratio of 25 was added to the time series data of the high consumption case to simulate common measurement noise affecting building measurements. Recent work by the same authors has shown the Bayesian inference approach to be robust with respect to a wide range of signal-to-noise ratios and noise colors [152].

Table 6.9 shows the nominal values used to generate the prior input parameter distributions as well as the parameter values associated with the low, medium, and high cases investigated in this section. The latter represent the ground truth values that will be compared to the posterior distributions for each model parameter to determine whether the measured data are used to update our belief of the uncertain parameters in a way that is consistent with the ground truth data.

The table shows that several individual parameters appear inconsistent with the case they belong to. As one example, in the low consumption case, the DX coil-rated COP is 3.1, which is below the nominal value. One would expect a higher COP to be associated with the low consumption case. However, given the tradeoffs between lighting and equipment power consumption and the efficiency of the cooling equipment, the low consumption case, at the 5th percentile of 10,000 simulation runs, was the result of a less efficient RTU with strongly reduced lighting and equipment power densities. The median consumption case resulted from a slightly higher lighting power density, lower equipment power density, lower occupant density, and higher RTU COP, all relative to the nominal values. The high consumption case, at the 95th percentile of all cases, is characterized by higher lighting, equipment, and occupancy densities, and lower HVAC equipment efficiency. In this case of higher consumption, the individual parameter values are all consistent with the theme of the case.

Table 6.9: Building Model Truth Parameters.

Parameter	Nominal	Low	Medium	High	Units
Lighting power density	32.30	26.99	33.09	36.15	W/m^2
Equipment power density	5.23	4.92	4.96	5.94	W/m^2
Occupant density	0.141	0.149	0.137	0.151	per/m^2
DX coil rated COP	3.20	3.10	3.39	2.94	-
Gas heating coil efficiency	80	74	70	77	%

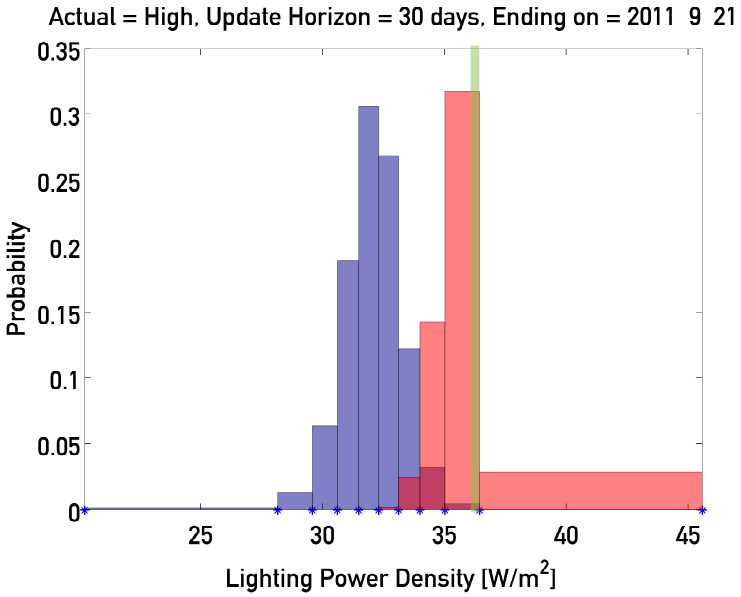

Figure 6.32: Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line.

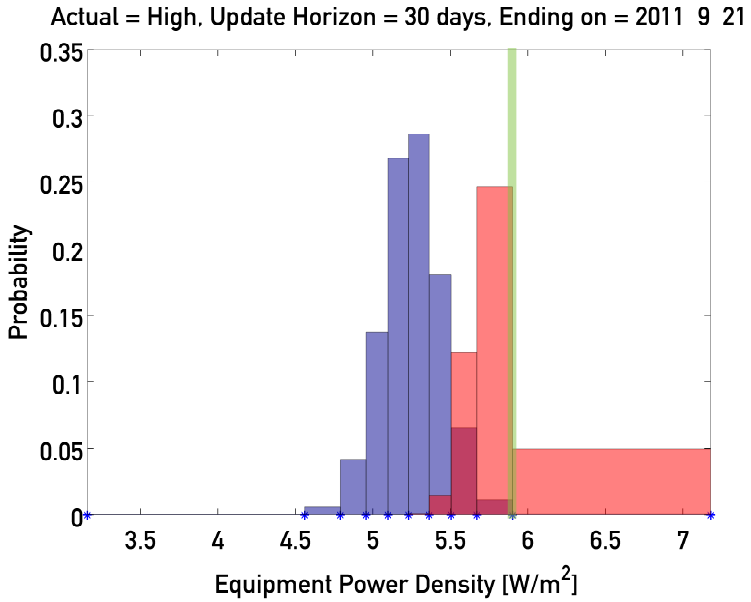

Figure 6.33: Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line.

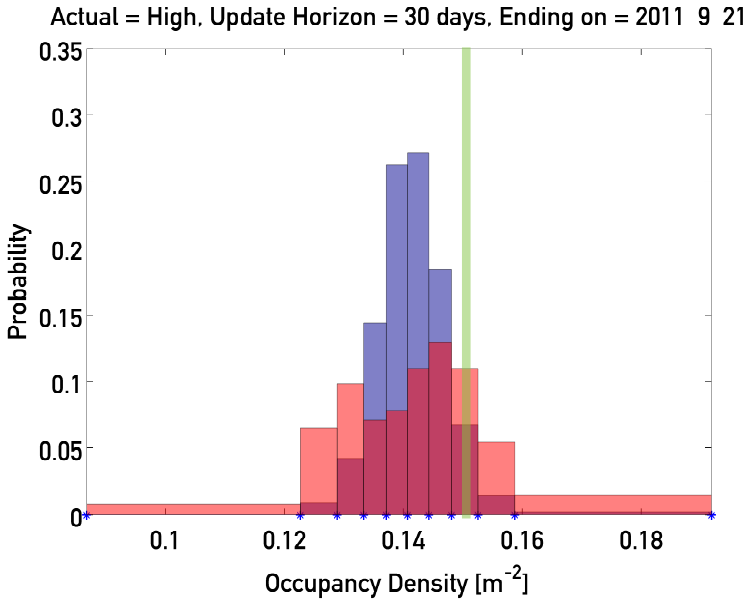

Figure 6.34: Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line.

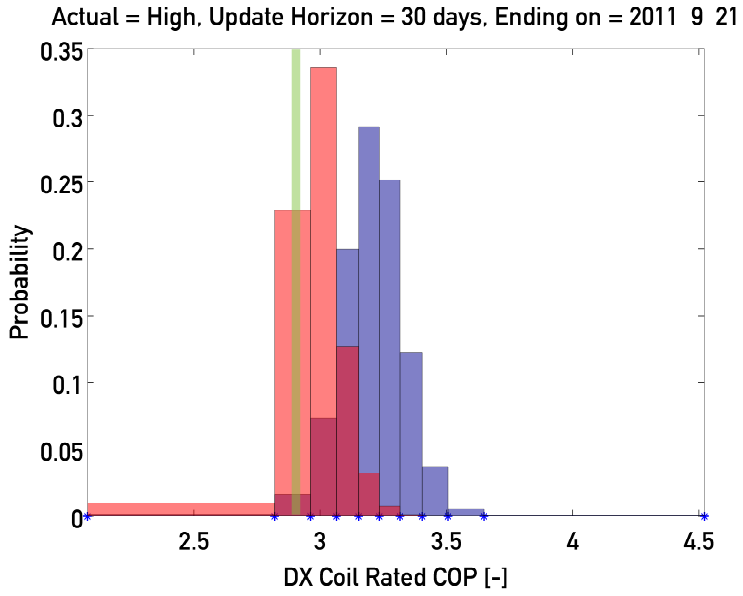

Figure 6.35: DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line.

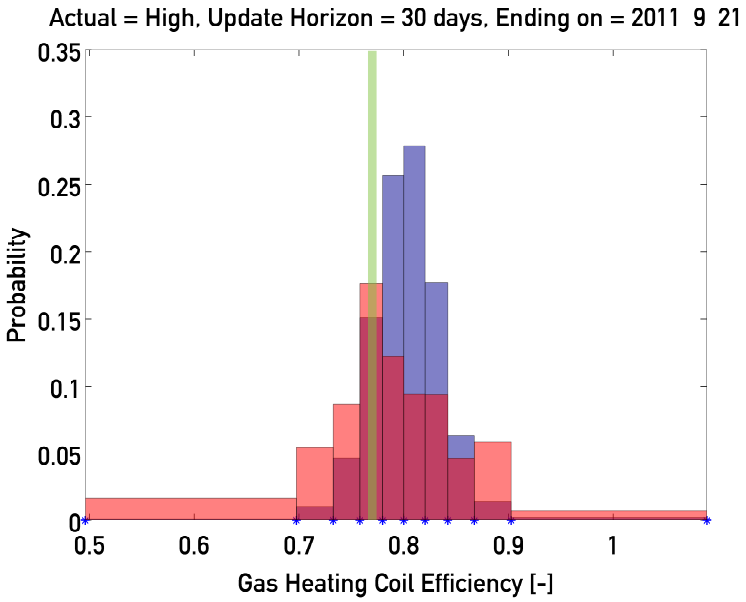

Figure 6.36: Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of high energy consumption. Truth value is shown as a green vertical line.

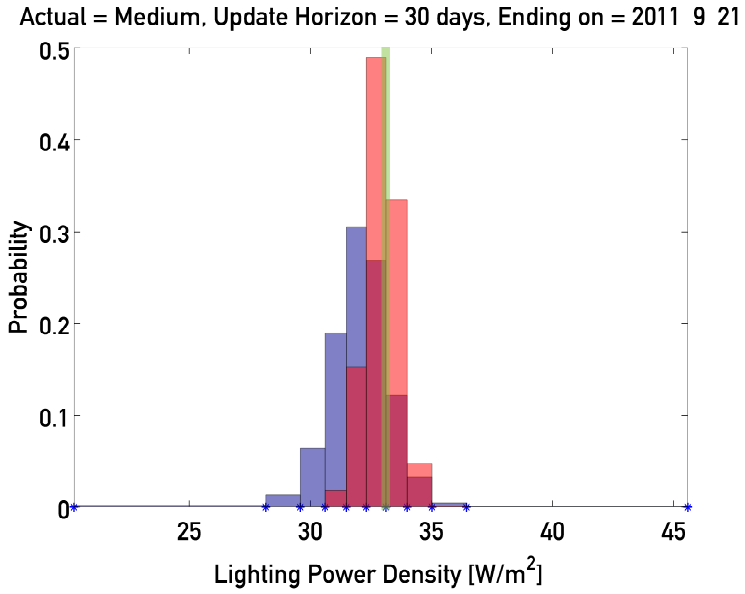

Figure 6.37: Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line.

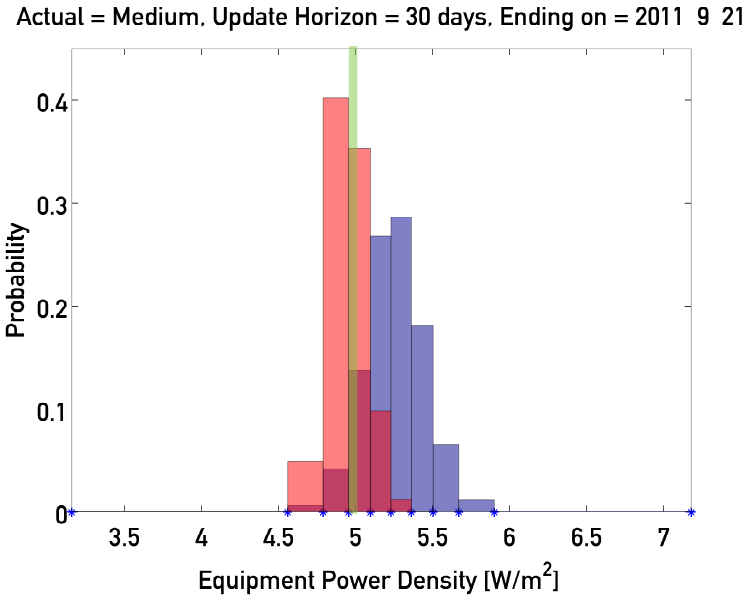

Figure 6.38: Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line.

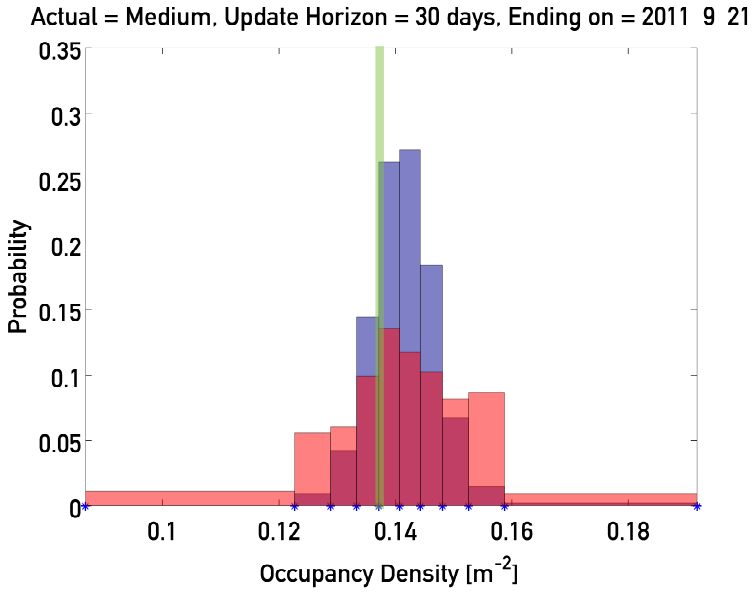

Figure 6.39: Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line.

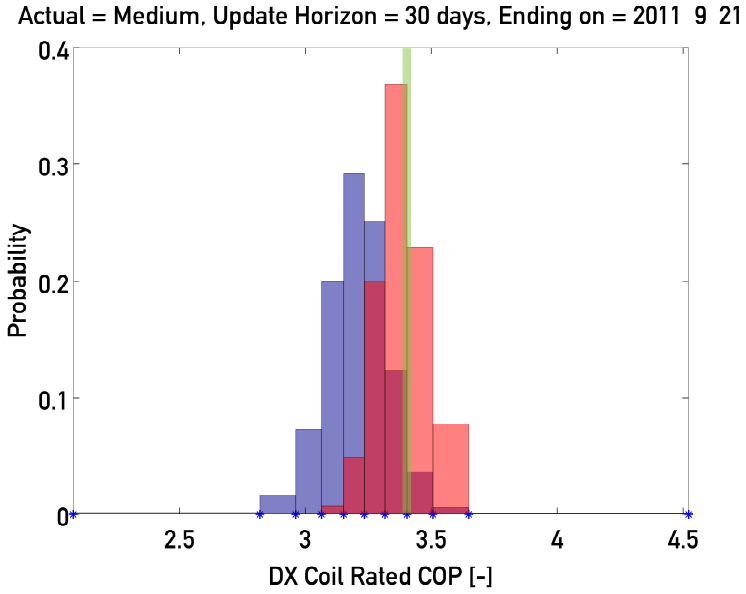

Figure 6.40: DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line.

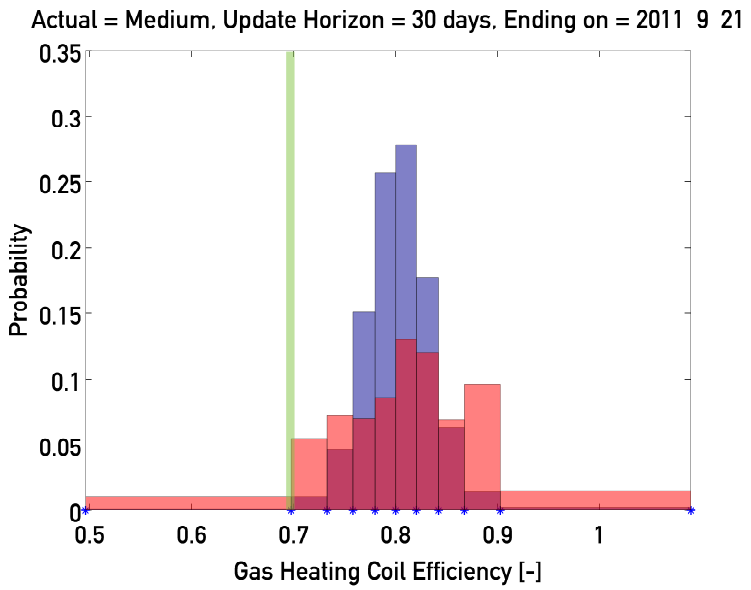

Figure 6.41: Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of medium energy consumption. Truth value is shown as a green vertical line.

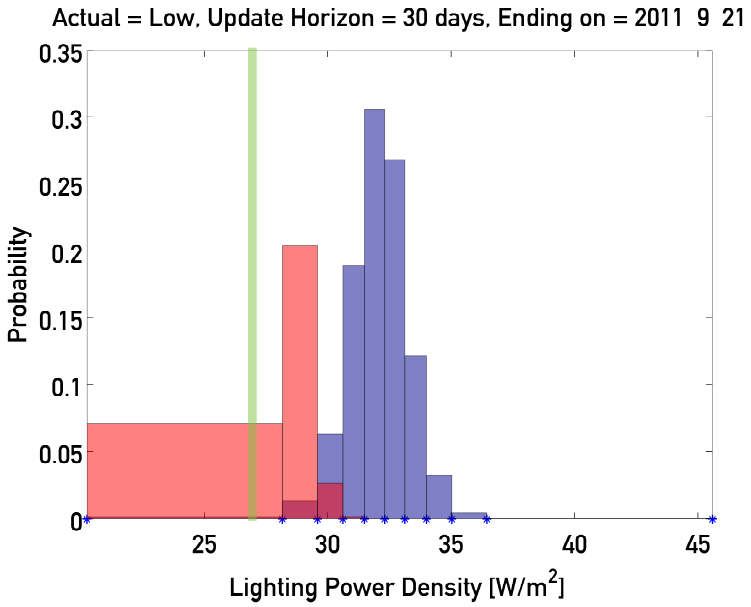

Figure 6.42: Lighting power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line.

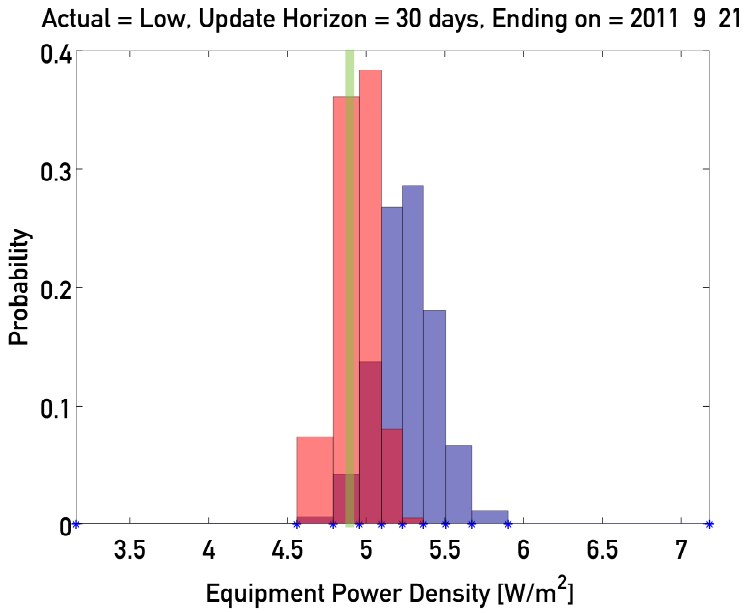

Figure 6.43: Equipment power density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line.

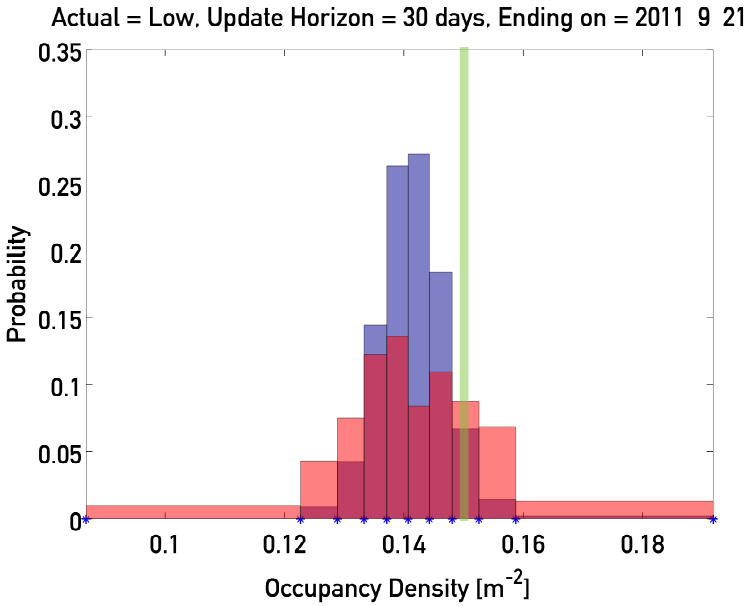

Figure 6.44: Occupant density prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line.

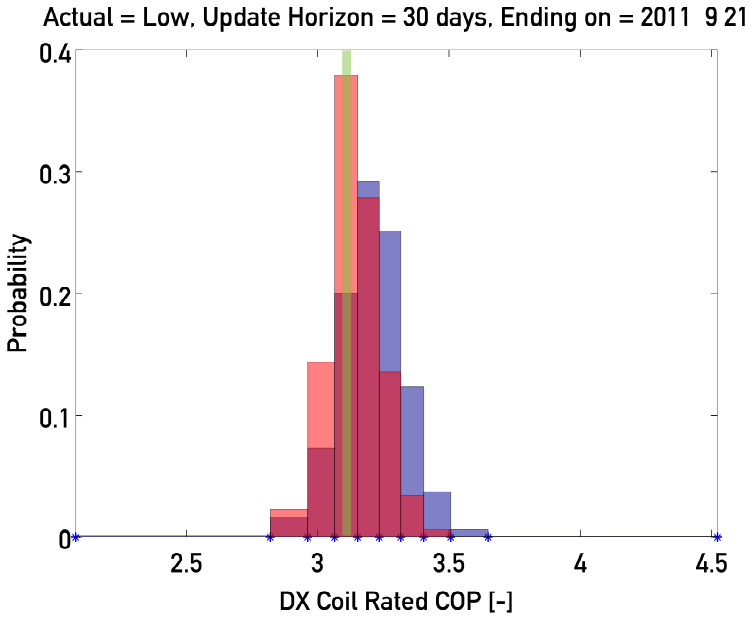

Figure 6.45: DX Coil Rated COP prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line.

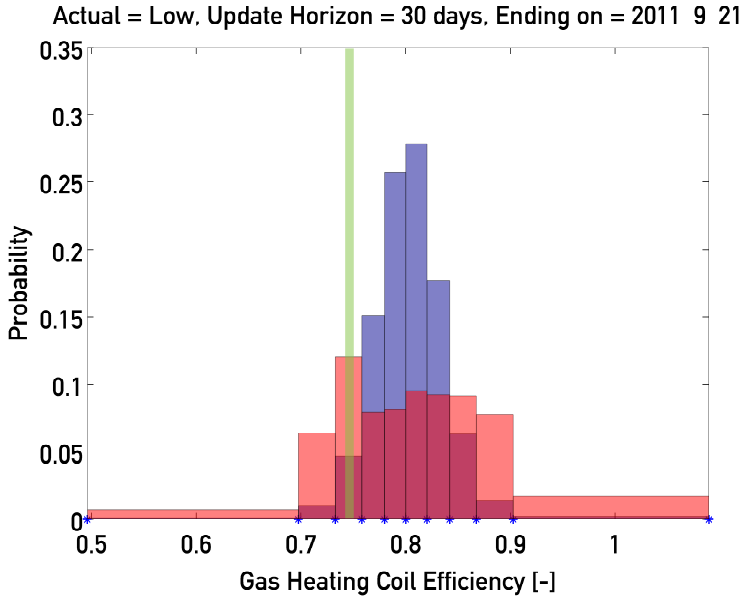

Figure 6.46: Gas heating coil efficiency prior (blue) and posterior (red) distributions on September 21 based on past 30 days of low energy consumption. Truth value is shown as a green vertical line.

Later sections discuss that whatever tradeoffs were at play in leading to the low, medium,

and high cases, the Bayesian parameter updating process is changing the posterior parameter distributions toward the ground truth values that form the basis of the measured (here surrogate) consumption data. To that effect, each of the following figures 6.32 to 6.46 shows the sampled prior distribution of the uncertain parameter as blue bars, the sampled posterior distribution as red bars, and the ground truth value that formed the basis of the surrogate measured data as a green vertical line.

Gas heating efficiency should have no impact on electricity consumption; thus, increased electricity consumption should be independent from gas heating coil efficiency. All the figures showing gas heating efficiency distributions illustrate that the gas heating efficiency posterior distribution is being smeared out; i.e., becoming less informative, because the evidence used in the likelihood function does not offer any clear clues about how to shape the posterior. Although the prior and posterior distributions for gas heating efficiency for all three cases are shown, the independence of electricity consumption from gas heating efficiency is seen in a widening posterior in each case.

6.4.2.1 High Energy Consumption Case

Figures 6.32 through 6.36 clearly illustrate how the measured data would be harnessed to update our belief about the uncertain parameters. Each figure shows the empirical prior distribution as blue bars; the empirical posterior distribution is shown as red bars. Where the two distributions overlap, a darker, purple hue appears. As stated above, in the high energy consumption case, all individual building parameter truth values are consistent with the theme of high energy consumption. Lighting and equipment power density posteriors move to higher values relative to the nominal values that served as the mean of the normal prior distributions and gravitate toward the truth values (green vertical lines) used to generate the surrogate measured data. Similarly, occupancy densities in people per square meter are also slightly higher and the DX coil-rated COP is significantly lower than the nominal value, gravitating toward the truth values. Thus, the Bayesian inference “learns” the truth values that form the basis of the measured data.

6.4.2.2 Medium Energy Consumption Case

Figures 6.37 through 6.41 show how in the medium consumption case, the data reveal the tradeoffs that form the basis of the medium consumption case: A slightly higher lighting power density is compensated for by a significantly lower equipment power density, because these two have the identical effects of adding convective internal gains to the sensible energy balance, paired with slightly higher COPs. Occupant density distribution has not materially changed from prior to posterior in the medium consumption case. As in the high consumption case, the posterior distributions (except for gas heating efficiency, as explained above) have moved in the direction of the truth values that form the foundation of the surrogate measured data.

6.4.2.3 Low Energy Consumption Case

Figures 6.42 through 6.46 again clearly illustrate how the measured data would be harnessed to update our belief about the uncertain parameters. In the low consumption case, the data suggest significantly lower lighting and equipment power densities, although occupant density seems to materially impact the measured data, which lead to a virtually unchanged posterior.

As shown in Table 6.9, the truth value of the DX coil-rated COP is slightly lower than the nominal value, which is opposite to the theme of lower energy consumption. As explained above, the slightly inferior COP is more than compensated for by significantly lower lighting and equipment power densities. The posterior COP distribution is close[3] to the prior but slightly less than it, feels like something is missing here suggesting that the Bayesian inference has "learned" the truth value.

As before, electricity consumption should not and does not have informative power for gas heating efficiency, leaving the posterior smeared out relative to the prior.

[3] A very heuristic, visual interpretation of "close" was used. The reader is simply encouraged to determine in which direction the posterior distribution mass has been moving: to the left, to the right, or virtually unchanged relative to the prior distribution. In particular, the point of using such loose language is to avoid comparing the empirical prior and posterior distributions with a more rigorous metric such as the maximum distance of the two cumulative distribution functions as used in the two-sample Kolmogorov-Smirnof test. Future work will look into more rigorous metrics to automate the update process. The point to make in this work, however, is that evidence is collected and used to inform updates of the input parameter distributions.

6.5 Summary and Conclusions

A prototype energy signal tool was demonstrated for operational whole-building and system-level energy performance assessment. The purpose of the tool is to give an assessment that a building operator or other user can quickly comprehend. Toward this end, the energy signal tool estimates energy use for various end uses from a low-order lumped-parameter model, taking into account uncertainty (via a Monte Carlo method) in model parameters and inputs. The result of the modeling phase is a probability distribution over estimated energy use. The range of estimated energy use is divided into intervals based on the observed energy use, and the probability that energy use is in an interval is computed as the mass of the estimated energy use distribution in that interval. An indicator (traffic light color) is chosen to minimize misclassification cost. Model parameter distributions are adjusted over time via Bayesian updating.

The experimental study investigated whole-building energy signal accuracy in the presence of uncertainty and faults at the submetered level, which may lead to tradeoffs at the whole-building level that are not detectable without submetering. Submetering of end uses is recommended to avoid confounding underconsumption and overconsumption among various end uses. An example of four building energy signal displays is offered to illustrate energy performance features that could be detected by the energy signal tool. The Bayesian inference results presented show that observations can be used to periodically update model parameter distributions and that the posterior distributions indeed gravitate toward the ground truth parameter values that formed the basis of the surrogate measured data. Results for a 30-day learning cycle are presented .

Future improvements in the inference process would eliminate the assumptions of temporal independence of subsequent observations of a particular variable and structural independence of multiple observed variables. Accounting for covariance among observed variables will help to better attribute observations to individual model parameters.

Chapter 7

Conclusions and Future Research

This chapter summarizes the high-level outcomes from research into *a data-driven toolchain for the operational performance analysis and optimization of buildings.* Due to its cumulative nature, individual contributions have been provided on a chapter-by-chapter basis. Chapters 4–6 form the main components of the data-driven toolchain by addressing classification, forecasting, and signaling, respectively. Combining these process components, the APEX system concept is perceived as the terminal decision support application; however, complications arise from two sources: 1) whether the iterative process will work for every one-of-a-kind building, and 2) whether the iterative process converges on a suitable OpBEM for prediction and diagnosis? Research outcomes showed promise but a complete software and hardware solution was beyond the scope of research. In addition, as evident from results of the ModBen project (Appendix A), automation is helpful but human experts are irreplaceable. The following sections discuss the big picture implications of the research.

7.1 Overview

It was determined that operational building energy performance can be conceived as a combination of prediction and diagnosis tasks for the purposes of decision support involving a building's HITL processes. The subjective nature of how an ideal building's performance can be defined remains an open question. A scalable and automated abstraction of operational building energy performance was therefore sought. A three-step process was proposed for the continual improve-

ment of building operations; i.e., the APEX system was posed as a concept to facilitate data-driven developments that are impactful, transferable, and meaningful to the smart grid and end-users alike. *Classification*, the first step of the process, can be thought of as the best estimate a human expert can obtain on a particular query using best-in-class, data-driven physical models. The second step in the process, *Forecasting*, makes it clear that planning over any horizon of interest involves uncertainties (i.e. forecast versus actual) in operational performance; any decision period of interest, whether day-ahead forecasting for weather-driven thermal loads or intra-day rooftop PV generation, inherently involves decisions in a sequential fashion and builds evidence over time. *Signaling*, the third step in the process, is retrospective in the sense of trying to seek and understand the errors, which are expected in one sense (e.g., weather forecasts), but are primarily used to *reduce uncertainty through learning* while inferring possible faulty behavior. The three-step process combines to provide decision support for buildings, and it is envisioned to be iterative. The following paragraphs summarize the process steps as applied to particular engineering problems.

Classification: Commercial buildings have a significant impact on energy and the environment, being responsible for more than 18% of the annual primary energy consumption in the United States. Analyzing their electrical demand profiles is necessary for the assessment of supply-demand interactions and potential; of particular importance are supply- or demand-side energy storage assets and the value they bring to various stakeholders in the smart grid context. This research developed and applied unsupervised classification of commercial buildings according to their electrical demand profile. A Department of Energy (DOE) database was employed, containing electrical demand profiles representing the United States commercial building stock as detailed in the 2003 Commercial Buildings Energy Consumption Survey (CBECS) and as modeled in the EnergyPlus building energy simulation tool. The essence of the approach was: 1) discrete wavelet transformation of the electrical demand profiles, 2) energy and entropy feature extraction (absolute and relative) from the wavelet levels at definitive time frames, and 3) Bayesian probabilistic hierarchical clustering of the features to classify the buildings in terms of similar patterns of electrical demand. The process yielded a categorized and more manageable set of representative electrical

demand profiles, inference of the characteristics influencing supply-demand interactions, and a test bed for quantifying the impact of applying energy storage technologies.

Forecasting: Model predictive control applied to commercial buildings requires short-term weather forecasts to optimally adjust setpoints in a supervisory control environment. Review of the literature reveals that many researchers are convinced that nonlinear forecasting models based on neural networks (NN) provide superior performance over traditional time series analysis. This research seeks to identify the complexity required for short-term weather forecasting in the context of a model predictive control environment. Moving average (MA) models with various enhancements and neural network models are used to predict weather variables seasonally in numerous geographic locations. Their performance is statistically assessed using coefficient-of-variation (CV) and mean bias error (MBE) values. When used in a cyclical two-stage model predictive control process of policy planning followed by execution, the results show that even the most complicated nonlinear autoregressive neural network with exogenous input (NARX) does not appear to warrant the additional efforts in forecasting model development and training in comparison to the simpler MA models.

Signaling: This project demonstrates a prototype energy signal tool for operational whole-building and system-level energy use evaluation. The purpose of the tool is to provide a summary of building energy use that allows a building operator to quickly distinguish normal and abnormal energy use. Toward that end, energy use status is displayed as a traffic light, which is a visual metaphor for energy use that is either substantially different from expected (red and yellow lights) or approximately the same as expected (green light). Which light to display for a given energy end use is determined by comparing expected to actual energy use. As expected, energy use is necessarily uncertain; we cannot choose the appropriate light with certainty. Instead, the energy signal tool chooses the light by minimizing the expected cost of displaying the wrong light. The expected energy use is represented by a probability distribution. Energy use is modeled by a low-order lumped parameter model. Uncertainty in energy use is quantified by a Monte Carlo exploration of the influence of model parameters on energy use. Distributions over model parameters are updated

over time via Bayes' theorem. The simulation study was devised to assess whole-building energy signal accuracy in the presence of uncertainty and faults at the submetered level, which may lead to tradeoffs at the whole-building level that are not detectable without submetering.

Next, the anticipated difficulties in practical application of the research are considered to motivate future R&D.

7.2 Uncertainty Quantification

Any model, whether physical, statistical, or some hybrid, can be feasibly considered for use as an OpBEM; however, uncertainties will always exist in one form or another. The below subsections explore the implications of this fact with regard to its preceived influence on the APEX system concept in practice.

Two Extrema for Learning The research takes the premise there are theoretical extrema regarding the task of learning causality; the minima being no data and no physical model, and the maxima being an infinite amount of data coupled to an infinitely detailed physical model. What can be accomplished, say, using brute force sampling approaches and physical simulation, for different data-model resolution between these extrema? At what point can one learn causality? It is argued the existing set of measurement data found in modern commercial buildings is sufficient for learning causality because they were specifically designed with building system control in mind. In addition, when appropriately-detailed physical models are available it is possible to diagnose faults influencing whole-building energy performance; i.e., latent variables that are part of the physical model but not the measurement dataset. However, although these statements seem valid, the infrequency by which physical models matched empirical time series data (ModBen discussion of Appendix A) could lead one to believe the signal-to-noise ratio was too small to maintain faith in the accuracy of the physical model. On the contrary, these buildings may be too coupled to the ambient (microclimate) to benefit from physical modeling for prediction and diagnosis. It is thus argued that greater focus should be placed on built environment R&D: one that makes progress from *data-to- models* and less focus on *models-to-(simulated)data.*

Ideal Building Energy Performance Ideal operational building energy performance can be specified according to a building energy performance rating/scoring program, model-based benchmark, or similar measures. It is assumed there exists some building energy simulation model that establishes an ideal energy performance. Any deviations from that ideal are considered faulty and symptomatic of a causal relationship to be inferred. That is, in fault diagnostics the concept of a 'fault' is phrased in terms of any deviation from the ideal. The prevalence of energy performance codes and standards, as well as model-based approaches for energy performance simulation and prediction, allows the ideal, and what constitutes faulty building energy performance, to be precisely defined. Measurement uncertainty and the subjective nature of metrics for one-of-a-kind buildings, however, do not allow straightforward application of these components to serve as a mechanism for energy performance guidance. Additional R&D is necessary to establish databases of nominal and fault behavior for one-of-a-kind buildings. It is envisioned this will take the form of some risk minimization simulation rather than worst-case deterministic evaluations.

Shades of Gray in Physical Modeling It is apparent there are many "shades of gray" between a pure, first principle model and a completely empirical one. Authors involved in building energy performance modeling [188, 28] distinguish between white box, gray box, and black box approaches. The categorization has been present in the literature for years, with a detailed account provided by Karplus [205], and briefly summarized as follows.

White Box Models are completely derived from first principles. Theoretical modeling allows all equations and variables to be defined mathematically, either implicitly or explicitly. However, due to practical considerations even most white models contain some parameters that are estimated from data. Even so, these parameters should be directly related to first principle modeling.

Black Box Models are completely derived from measured data. The structure and parameters of the model are completely informed from experiments and/or other measurement campaigns. Because of the automation available in the computer age, very little prior knowledge is

necessary to implement a black box model. Furthermore, because of the numerical nature of the model, its direct relationship to physical observation is severely limited.

Gray Box Models are a mixture, combination, or compromise between black box and white box model. Differences between the various shades of gray box models can be attributed to the validation stage of the system identification process, where adjustments are made to enhance the accuracy of model. It is typical that the model structure is strongly driven by prior knowledge, whereas parameter values are primarily determined from measure data.

As one can imagine there are practically no purely white box or black box models. That is, for realistic, complex models there are always subjective aspects inherent to the modeling process (i.e., no completely automated system identification), and some data are inevitably used even for equation-based model building (e.g., properties of materials). The shade of gray employed also implies understanding of the physical mechanisms important to the problem; i.e., a detailed simulation is overkill if a linear regression on experimental data answers the question posed. In any case, aspects of semi-physical (gray box) modeling are helpful, and in accordance with the scientific method, for using both first principles and measured data in two crucial ways. The first is to validate models, or replicate results, from seemly disparate research. The second is to condition analysis, or further research, on state of the art models and understanding of physical processes. In addressing the energy concerns of the built environment, movement must be made away from simulated energy performance and toward measured energy performance.

7.3 Future Research

In summary, the iterative, three-step process of Figure 3.2 shows there is much to be done with regard to the classification, forecasting, and signaling of building energy performance; i.e., the research has made it one time around the spiral graph. Future work will look define the OpBEM in greater detail and implement it within the APEX system concept for uncountable HITL objectives.

Bibliography

[1] Le Corbusier. Toward an architecture. Getty Publications, 2007.

[2] K. Ashton. That 'internet of things' thing. RFiD Journal, 22(7):97–114, 6 2009.

[3] U.S. Department of Energy. Annual energy review 2011. Technical report, U.S. Energy Information Administration, Washington, D.C., 9 2012.

[4] U.S. Department of Energy. 2011 buildings energy data book. Technical report, Energy Efficiency & Renewable Energy, Washington, D.C., 3 2012.

[5] S. Katipamula and M.R. Brambley. Review article: methods for fault detection, diagnostics, and prognostics for building systems - a review, part i. HVAC&R Research, 11(1):3–25, 1 2005.

[6] S. Katipamula and M.R. Brambley. Review article: methods for fault detection, diagnostics, and prognostics for building systems - a review, part ii. HVAC&R Research, 11(2):169–187, 4 2005.

[7] M. Mather, K.M. Pollard, and L. Jacobsen. First results from the 2010 census. Technical report, Population Reference Bureau, Washington, D.C., 7 2011.

[8] B. Williams. Heavy hydrocarbons playing key role in peak-oil debate, future energy supply. Oil & Gas Journal, 101(29):20–27, 7 2003.

[9] S. Solomon, G.-K. Plattner, R. Knutti, and P. Friedlingstein. Irreversible climate change due to carbon dioxide emissions. Proceedings of the national academy of sciences of the USA, 106(6):1704–1709, 2 2009.

[10] B. Rajagopalan, K. Nowak, J. Prairie, M. Hoerling, B. Harding, J. Barsugli, A. Ray, and B. Udall. Water supply risk on the colorado river: Can management mitigate? Water Resources Research, 45, 8 2009.

[11] M.A. Maupin, J.F. Kenny, S.S. Hutson, J.K. Lovelace, N.L. Barber, and K.S. Linsey. Estimated use of water in the united states in 2010. Technical report, U.S. Geological Survey, 11 2014.

[12] U.S. Department of Energy. Annual energy outlook 2012. Technical report, U.S. Energy Information Administration, Washington, D.C., 6 2012.

[13] A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D.E. Culler, and R.H. Katz. Integrating renewable energy using data analytics systems: challenges and opportunities. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 34(1):3–11, 3 2011.

[14] U.S. Department of Energy. International energy outlook 2011. Technical report, U.S. Energy Information Administration, Washington, D.C., 9 2011.

[15] K. Gillingham, D. Rapson, and G. Wagner. The rebound effect and energy efficiency policy. Review of Environmental Economics and Policy, 10(1):68–88, 2016.

[16] C. Neumann, D. Jacob, S. Burhenne, A. Florita, E. Burger, and F. Schmidt. Model-based methods for fault detection and optimization in building operations (modellbasierte methoden fur die fehlererkennung und optimierung im gebaudebetrieb). Technical report, Fraunhofer ISE, 2011.

[17] J. Hirshleifer and J.G. Riley. The analytics of uncertainty and information-an expository survey. Journal of Economic Literature, 17(4):1375–1421, 12 1979.

[18] R.H. Dodier. Unified prediction and diagnosis in engineering systems by means of distributed belief networks. PhD thesis, University of Colorado Boulder, 1999.

[19] J.J. Romm and W.D. Browning. Greening the building and the bottom line: Increasing productivity through energy-efficient design. Technical report, Rocky Mountain Institute, 1994.

[20] A. Florita, B.-M. Hodge, and K. Orwig. Identifying wind and solar ramping events. In 2013 IEEE Green Technologies Conference (GreenTech), pages 147–152. IEEE, 2013.

[21] W.A. Shewhart and W.E. Deming. Statistical method from the viewpoint of quality control. Department of Agriculture, Washington, D.C., 1939.

[22] C. Eastman, P. Teicholz, R. Sacks, and K. Liston. BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors. Wiley, 2011.

[23] J. Clarke. Energy simulation in building design. Routledge, 2012.

[24] T. Kusuda. Early history and future prospects of building system simulation. In Proceedings of Building Simulation'99, volume 1, pages 3–16, 1999.

[25] C. Underwood and F. Yik. Modelling methods for energy in buildings. Wiley-Blackwell, 2008.

[26] J. Neymark, R. Judkoff, G. Knabe, H.-T. Le, M. Durig, A. Glass, and G. Zweifel. Applying the building energy simulation test (bestest) diagnostic method to verification of space conditioning equipment models used in whole-building energy simulation programs. Energy and Buildings, 34(9):917–931, 2002.

[27] A. Malkawi and G. Augenbroe. Advanced building simulation. Taylor & Francis, 2004.

[28] J.L.M. Hensen and R. Lamberts. Building performance simulation for design and operation. Routledge, 2011.

[29] I.A. Macdonald. Quantifying the effects of uncertainty in building simulation. PhD thesis, University of Strathclyde, 2002.

[30] A. Ahmed, J. Ploennigs, K. Menzel, and B. Cahill. Multi-dimensional building performance data management for continuous commissioning. Advanced Engineering Informatics, 24(4):466–475, 2010.

[31] M. Way and B. Bordass. Making feedback and post-occupancy evaluation routine 2: Soft landings–involving design and building teams in improving performance. Building Research & Information, 33(4):353–360, 2005.

[32] G.P. Henze, C. Felsmann, A.R. Florita, M.J. Brandemuehl, H. Cheng, and C.E. Waters. Optimization of building thermal mass control in the presence of energy and demand charges (rp-1313). ASHRAE Transactions, 114(2):75, 2008.

[33] R.Y. Rubinstein and D.P. Kroese. Simulation and the Monte Carlo method, volume 707. Wiley-interscience, 2011.

[34] T. A. Reddy and I. Maor. Procedures for Reconciling Computer-Calculated Results With Measured Energy Data, ASHRAE Research Project 1051-RP. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA, 2006.

[35] R.H. Dodier, P.S. Curtiss, and J.F. Kreider. Small-scale on-line diagnostics for an hvac system. Transactions-American Society Of Heating Refrigerating And Air Conditioning Engineers, 104:530–539, 1998.

[36] R.H. Dodier and J.F. Kreider. Whole building energy diagnostics. ASHRAE Trans, 105(1):579–589, 1999.

[37] S. Kaldorf, P. Gruber, and G. Wichenko. Practical experiences from developing and implementing an expert system diagnostic tool. ASHRAE Transactions, 108 PART 1:826 – 840, 2002. Building energy management systems (BEMS);.

[38] L. Song, M. Liu, D.E. Claridge, and P. Haves. Study of on-line simulation for whole building level energy consumption fault detection and optimization. In Architectural Engineering, pages 76 – 83, Austin, TX, USA, 2003.

[39] G.E. Metcalf and K.A. Hassett. Measuring the energy savings from home improvement investments: Evidence from monthly billing data. Review of economics and statistics, 81(3):516–528, 1999.

[40] R.E. Mortensen and K.P. Haggerty. Dynamics of heating and cooling loads: Models, simulation, and actual utility data. Power Systems, IEEE Transactions on, 5(1):243–249, 1990.

[41] M. Berges, E. Goldman, H.S. Matthews, and L. Soibelman. Training load monitoring algorithms on highly sub-metered home electricity consumption data. Tsinghua Science & Technology, 13:406–411, 2008.

[42] W. Kastner, G. Neugschwandtner, S. Soucek, and H.M. Newmann. Communication systems for building automation and control. Proceedings of the IEEE, 93(6):1178–1203, 2005.

[43] ASHRAE. Guideline 14-2014. Measurement of Energy, Demand, and Water Savings, 2014.

[44] S. Bertagnolio, V. Lemort, and P. Andre. Simulation assisted audit & evidence based calibration methodology. Technical report, Thermodynamics Laboratory, University of Liége, 2010.

[45] P. Raftery, M. Keane, and J. O'Donnell. Calibrating whole building energy models: An evidence-based methodology. Energ. Buildings, 43:2356–2364, 2011.

[46] L. K. Norford, R. H. Socolow, E. S. Hsieh, and G. V. Spadaro. Two-to-one discrepancy between measured and predicted performance of a 'low-energy' office building: insights from a reconciliation based on the doe-2 model. Energy and Buildings Buildings, 21:121–131, 1994.

[47] S. U. Lee and D. E. Claridge. Automatic calibration of a building energy simulation model using a global optimization program. In Proceedings of the Second International Conference for Enhanced Building Operations, Richardson, TX, 2002, 2002.

[48] A. Pedrini, F. S. Westphal, and R. Lamberts. A methodology for building energy modelling and calibration in warm climates. Build. Environ., 37:903–912, 2002.

[49] J. Yoon, E.J. Lee, and D. E. Claridge. Calibration procedure for energy performance simulation of a commercial building. J. Sol. Energ., 125:251–257, 8 2003.

[50] B. Eisenhower, Z. O'Neill, S. Narayanan, V.A. Fonoberov, and I. Mezić. A methodology for meta-model based optimization in building energy models. Energy and Buildings, 47:292–301, 2012.

[51] B. Eisenhower, Z. O'Neill, V.A. Fonoberov, and I. Mezić. Uncertainty and sensitivity decomposition of building energy models. Journal of Building Performance Simulation, 5(3):171–184, 2012.

[52] S. Burhenne, O. Tsvetkova, D. Jacob, G.P. Henze, and A. Wagner. Uncertainty quantification for combined building performance and cost-benefit analyses. Building and Environment, 62:143–154, 2013.

[53] U.S. Department of Energy. 2003 commercial buildings energy consumption survey. Technical report, U.S. Energy Information Administration, Washington, D.C., 2005.

[54] U.S. Department of Energy. 2005 residential energy consumption survey. Technical report, U.S. Energy Information Administration, Washington, D.C., 2009.

[55] L. Ljung. System identification. John Wiley & Sons, Inc., 1999.

[56] O. Nelles. Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer, 2000.

[57] R. Pintelon and J. Schoukens. System identification: a frequency domain approach. Wiley-IEEE Press, 2004.

[58] ASHRAE. Handbook of Fundamentals. American Society of Heating, Refrigerating and Air Conditioning Engineers, 2009.

[59] G.S. Pavlak, G.P. Henze, and V.J. Cushing. Optimizing commercial building participation in energy and ancillary service markets. Energy and Buildings, 81:115–126, 2014.

[60] N.F. Saman, A.J. Britton, T.A. Reddy, and W.M. Heffington. Empirical correlations between annual cost savings and implementation costs for various categories of energy conservation retrofit projects in commercial buildings. In International Solar Energy Conference, volume 1, pages 135 – 144, Maui, HI, USA, 1995.

[61] M. Bauer and J.-L. Scartezzini. Simplified correlation method accounting for heating and cooling loads in energy-efficient buildings. Energy and Buildings, 27(2):147 – 154, 1998.

[62] S. Hammarsten. Critical appraisal of energy-signature models. Applied Energy, 26(2):97 – 110, 1987.

[63] R.C. Sonderegger. Baseline model for utility bill analysis using both weather and non-weather-related variables. ASHRAE Transactions, 104(2):859 – 870, 1998. Baseline equation;.

[64] A. Rabl and A. Rialhe. Energy signature models for commercial buildings: test with measured data and interpretation. Energy and Buildings, 19(2):143 – 154, 1992.

[65] S. Katipamula, R.T. Agami, and D.E. Claridge. Effect of time resolution on statistical modeling of cooling energy use in large commercial buildings. ASHRAE Transactions, 101(Pt 2):172 – 185, 1995.

[66] T.A. Reddy and D.E. Claridge. Uncertainty of 'measured' energy savings from statistical baseline models. HVAC and R Research, 6(1):3 – 20, 2000. Statistical baseline models;.

[67] S. Katipamula and J.S. Haberl. Methodology to identify diurnal load shapes for non-weather dependent electric end-uses. ASME-JSES-JSME International Solar Energy Conference, pages 457 – 467, 1991.

[68] D.L. Hadley. Daily variations in hvac system electrical energy consumption in response to different weather conditions. Energy and Buildings, 19(3):235 – 247, 1993.

[69] J.E. Seem. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. Energy and Buildings, 37(2):127 – 139, 2005.

[70] J.E. Seem. Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy and Buildings, 39(1):52 – 58, 2007.

[71] C. Miller, Z. Nagy, and A. Schlueter. Automated daily pattern filtering of measured building performance data. Automation in Construction, 49:1–17, 2015.

[72] R. Isermann. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer Science & Business Media, 2006.

[73] J. Hyvarinen and S. Karki, editors. IEA Annex 25 Real Time Simulation of HVAC Systems for Building Optimisation, Fault Detection and Diagnosis: Building Optimisation and Fault Diagnosis System Concept. International Energy Agency, 1993.

[74] P. B. Usoro, I. C. Schick, and S. Negahdaripour. Hvac system fault detection and diagnosis. In Proceedings of the American Control Conference, pages 606 – 612, Boston, MA, USA, 1985.

[75] J.E. Braun. Automated fault detection and diagnostics for vapor compression cooling equipment. Journal of Solar Energy Engineering, Transactions of the ASME, 125(3):266 – 274, 2003.

[76] P. Haves and S.K. Khalsa. Model-based performance monitoring: Review of diagnostic methods and chiller case study. Proceedings ACEEE Summer Study on Energy Efficiency in Buildings, 3:3161 – 3171, 2000. Fault detection;.

[77] F.W. Yu and K.T. Chan. Energy signatures for assessing the energy performance of chillers. Energy and Buildings, 37(7):739 – 746, 2005.

[78] K.K. Andersen and A.T. Reddy. The error in variables (eiv) regression approach as a means of identifying unbiased physical parameter estimates: Application to chiller performance data. HVAC&R Research, 8(3):295 – 310, 2002.

[79] S. Wang. Dynamic simulation of a building central chilling system and evaluation of emcs on-line control strategies. Building and Environment, 33(1):1 – 20, 1998. Central chilling systems;.

[80] S. Wang and J. Cui. A robust fault detection and diagnosis strategy for centrifugal chillers. HVAC and R Research, 12(3):407 – 428, 2006.

[81] Y. Jia and A.T. Reddy. Characteristic physical parameter approach to modeling chillers suitable for fault detection, diagnosis, and evaluation. Journal of Solar Energy Engineering, Transactions of the ASME, 125(3):258 – 265, 2003.

[82] Y. Jiang, Y. Yao, S. Deng, and Z. Ma. Applying grey forecasting to predicting the operating energy performance of air cooled water chillers. International Journal of Refrigeration, 27(4):385 – 392, 2004.

[83] N. S. Castro and G. Remington. Performance evaluation of a reciprocating chiller using experimental data and model predictions for fault detection and diagnosis. ASHRAE Transactions, 108 PART 1:889 – 903, 2002.

[84] M. Breuker, T. Rossi, and J.E. Braun. Smart maintenance for rooftop units. ASHRAE Journal, 42(11):41 – 42, 2000. Smart maintenance;Fault detection and diagnostic system;Rooftop air conditioners;.

[85] K. Choi, M. Namburu, M. Azam, J. Luo, K. Pattipati, and A. Patterson-Hine. Fault diagnosis in hvac chillers using data-driven techniques. In AUTOTESTCON, pages 407 – 413, San Antonio, TX, United States, 2004. Fault detection and isolation (FDI);Knowledge bases;Support vector machines (SVM);Diagnostic process;.

[86] S.M. Namburu, J. Luo, M. Azam, K. Choi, and K.R. Pattipati. Fault detection, diagnosis and data-driven modeling in hvac chillers. In Proceedings of SPIE - The International Society for Optical Engineering, volume 5809, pages 143 – 154, Orlando, FL, United States, 2005.

[87] S.M. Namburu, M.S. Azam, J. Luo, K. Choi, and K.R. Pattipati. Data-driven modeling, fault diagnosis and optimal sensor selection for hvac chillers. IEEE Transactions on Automation Science and Engineering, 4(3):469 – 473, 2007.

[88] X. Huang, H. Qi, and X. Liu. Implementation of fault detection and diagnosis system for control systems in thermal power plants. In Proceedings of the World Congress on Intelligent Control and Automation (WCICA), volume 2, pages 5777 – 5781, Dalian, China, 2006.

[89] W.-Y. Lee, C. Park, J.M. House, and G.E. Kelly. Fault diagnosis of an air-handling unit using artificial neural networks. ASHRAE Transactions, 102(1):540 – 549, 1996.

[90] W.-Y. Lee, J.M. House, and D.R. Shin. Fault diagnosis and temperature sensor recovery for an air-handling unit. ASHRAE Transactions, 103(1):621 – 633, 1997.

[91] W.-Y. Lee, J.M. House, and N.-H. Kyong. Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks. Applied Energy, 77(2):153 – 170, 2004.

[92] W.-Y. Lee, C. Park, and G.E. Kelly. Fault detection in an air-handling unit using residual and recursive parameter identification methods. ASHRAE Transactions, 102(1):528 – 539, 1996.

[93] S.H. Karki and S.J. Karjalainen. Performance factors as a basis of practical fault detection and diagnostic methods for air-handling units. ASHRAE Transactions, 105(pt 1):1069 – 1077, 1999. Fault detection and diagnostic methods;Air handling units;.

[94] B. Chen and J.E. Braun. Simple rule-based methods for fault detection and diagnostics applied to packaged air conditioners. ASHRAE Transactions, 107 PART 1:847 – 857, 2001. Automatic fault detection and diagnosis;Thermal expansion valves;Sensitivity ratios;.

[95] J. Schein and S.T. Bushby. Fault detection and diagnostics for ahus and vav boxes. ASHRAE Journal, 47(7):58 – 63, 2005.

[96] P. Xu, P. Haves, M. Kim, and B. Billedeaux. Model-based automated functional testing - methodology and application to air-handling units. ASHRAE Transactions, 111 PART 1:979 – 989, 2005.

[97] F. Xiao, S. Wang, and X. Xu. Automatic commissioning of ahu sensors for enhancing building lifecycle performance. Transactions Hong Kong Institution of Engineers, 15(1):24 – 29, 2008.

[98] S. Katipamula, R.G. Pratt, D.P. Chassin, Z.T. Taylor, K. Gowri, and M.R. Brambley. Automated fault detection and diagnostics for outdoor-air ventilation systems and economizers: methodology and results from field testing. ASHRAE Transactions, 105(pt 1):555 – 567, 1999.

[99] M. Liu, L. Song, G. Wei, and D.E. Claridge. Simplified building and air-handling unit model calibration and applications. Journal of Solar Energy Engineering, Transactions of the ASME, 126(1):601 – 609, 2004.

[100] S. Wang and F. Xiao. Sensor fault detection and diagnosis of air-handling units using a condition-based adaptive statistical method. HVAC and R Research, 12(1):127 – 150, 2006.

[101] J.E. Pakanen and T. Sundquist. Automation-assisted fault detection of an air-handling unit; implementing the method in a real building. Energy and Buildings, 35(2):193 – 202, 2003.

[102] J.M. House, K.D. Lee, and L.K. Norford. Controls and diagnostics for air distribution systems. Journal of Solar Energy Engineering, Transactions of the ASME, 125(3):310 – 317, 2003.

[103] J. Schein and S.T. Bushby. A hierarchical rule-based fault detection and diagnostic method for hvac systems. HVAC and R Research, 12(1):111 – 125, 2006. Fault detection and diagnostic (FDD) method;Rule based FDD;.

[104] J.E. Seem, J.M. House, and R.H. Monroe. On-line monitoring and fault detection. ASHRAE Journal, 41(7):21 – 26, 1999.

[105] C.C. Schroeder and J.D. Bradford. Model-independent fault detection and diagnostic for vav terminal units. Proceedings ACEEE Summer Study on Energy Efficiency in Buildings, 7:7 – 159, 2000.

[106] P. Fasolo and D.E. Seborg. Sqc approach to monitoring and fault detection in hvac control systems. In Proceedings of the American Control Conference, volume 3, pages 3055 – 3059, Baltimore, MD, USA, 1994. Statistical quality control;Shewhart chart;Chart violation;Building control systems;CUSUM chart;.

[107] P.S. Fasolo and D.E. Seborg. Monitoring and fault detection for an hvac control system. HVAC&R Research, 1(3):177 – 193, 1995. Fault detection;HVAC control system;Process variability;Air duct heating coil;.

[108] T. I. Salsbury and R. C. Diamond. Fault detection in hvac systems using model-based feedforward control. Energy and Buildings, 33(4):403 – 415, 2001.

[109] X. Hao, G. Zhang, and Y. Chen. Fault-tolerant control and data recovery in hvac monitoring system. Energy and Buildings, 37(2):175 – 180, 2005.

[110] A. Singhal and T.I. Salsbury. Characterization and cancellation of static nonlinearity in hvac systems. ASHRAE Transactions, 113 PART 1:391 – 399, 2007.

[111] T. Lewis. Don't let artificial intelligence take over, top scientists warn. Live Science, 2015.

[112] A.R. Florita, L.J. Brackney, T.P. Otanicar, and J. Robertson. Classification of commercial building electrical demand profiles for energy storage applications. Journal of Solar Energy Engineering, 135(3):031020, 2013.

[113] K.H. Lee and J.E. Braun. An Experimental Evaluation of Demand Limiting Using Building Thermal Mass in a Small Commercial Building. ASHRAE Transactions, 112(2):559–571, 2006.

[114] G.P. Henze, T. Le, A.R. Florita, and C. Felsmann. Sensitivity Analysis of Optimal Building Thermal Mass Control. Journal of Solar Energy Engineering, 129(4):473–485, 2007.

[115] W.P. Bahnfleth and W.S. Joyce. Energy Use in a District Cooling System with Stratified Chilled-Water Storage. ASHRAE Transactions, 100(1):1767–1778, 1994.

[116] C.W. Sohn, J. Fuchs, and M. Gruber. Chilled Water Storage Cooling System for an Army Installation. ASHRAE Transactions, 105(2):1126–1133, 1999.

[117] G.P. Henze. An Overview of Optimal Control for Central Cooling Plants with Ice Thermal Energy Storage. Journal of Solar Energy Engineering, 125(3):302–309, 2003.

[118] J.E. Braun. A Near-Optimal Control Strategy for Cool Storage Systems with Dynamic Electric Rates (RP-1252). HVAC&R Research, 13(4):557–580, 2007.

[119] A.J. Cavallo. Energy Storage Technologies for Utility Scale Intermittent Renewable Energy Systems. Journal of Solar Energy Engineering, 123(4):387–389, 2001.

[120] W. Kempton and J. Tomic. Vehicle-to-Grid Power Implementation: From Stabilizing the Grid to Supporting Large-Scale Renewable Energy. Journal of Power Sources, 144(1):280–294, 2005.

[121] U.S. Department of Energy. Open energy information: Commercial building profiles, 2010.

[122] U.S. Department of Energy. EnergyPlus Version 2.0, 2007.

[123] B. Griffith, N. Long, P. Torcellini, R. Judkoff, D. Crawley, and J. Ryan. Methodology for modeling building energy performance across the commercial sector. Technical report, National Renewable Energy Laboratory, Golden, CO, 2008.

[124] D.B. Percival and A.T. Walden. Wavelet Methods for Time Series Analysis. Cambridge University Press, 2000.

[125] S.G. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions of Pattern Analysis and Machine Intelligence, 11(7):674–693, 1989.

[126] K.A. Heller and Z. Ghahramani. Bayesian Hierarchical Clustering. In Proceedings of the 22nd International Conference on Machine Learning, pages 297–304, 2005.

[127] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2010.

[128] A.R. Florita and G.P. Henze. Comparison of short-term weather forecasting models for model predictive control. HVAC&R Research, 15(5):835–853, 2009.

[129] G.E.P Box, G.M. Jenkins, and G.C. Reinsel. Time Series Analysis: Forecasting & Control (3rd Edition). Prentice Hall, 1994.

[130] C.M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1995.

[131] J.W. MacArthur, A. Mathur, and J. Chao. On-line recursive estimation for load profile prediction. ASHRAE Transactions, 95(1):621–628, 1989.

[132] J.E. Seem and J.E. Braun. Adaptive methods for real-time forecasting of building electrical demand. ASHRAE Transactions, 97(1):710 – 721, 1991.

[133] T.Y. Chen and A.K. Athienitis. Ambient temperature and solar radiation prediction for predictive control of hvac systems and a methodology for optimal building heating dynamic operation. ASHRAE Transactions, 102(1):26 – 35, 1996.

[134] G.P. Henze, D.E. Kalz, C. Felsmann, and G. Knabe. Impact of forecasting accuracy on predictive optimal control of active and passive building thermal storage inventory. HVAC and R Research, 10(2):153 – 178, 2004.

[135] F. J. Ferrano and K.-F.V. Wong. Prediction of thermal storage loads using a neural network. ASHRAE Transactions, pages 723 – 726, 1990.

[136] J.F. Kreider and X.A. Wang. Improved artificial neural networks for commercial building energy use prediction. In Solar Engineering: Proceedings of Annual ASME Solar Energy Conference, pages 361 – 366, Maui, HI, USA, 1992.

[137] F.J. Gibson and T.T. Kraft. Electric demand prediction using artificial neural network techonology. ASHRAE Journal, 35(3):60–68, 1993.

[138] R.H. Dodier and G.P. Henze. Statistical analysis of neural networks as applied to building energy prediction. Journal of Solar Energy Engineering, 126(1):592–600, 2004.

[139] J.F. Kreider and J.S. Haberl. Predicting hourly building energy usage. ASHRAE Journal, 36(6):72 – 81, 1994.

[140] M. Kawashima, C.E. Dorgan, and J.W. Mitchell. Hourly thermal load prediction for the next 24 hours by arima, ewma, lr, and an artificial neural network. ASHRAE Transactions, 101(1):186 – 200, 1995.

[141] J.S. Haberl and S. Thamilseran. Great energy predictor shootout ii: measuring retrofit savings - overview and discussion of results. ASHRAE Transactions, 102(2):419 – 435, 1996.

[142] G.P. Henze, R.H. Dodier, and M. Krarti. Development of a predictive optimal controller for thermal energy storage systems. HVAC and R Research, 3(3):233–264, 1997.

[143] D.A. Rand and L.S. Young. Detecting strange attractors in turbulence. In F. Takens, editor, Dynamical Systems and Turbulence, Lecture Notes in Mathematics, pages 366–381. Berlin: Springer-Verlag, 1981.

[144] M. Hollander and D.A. Wolfe. Nonparametric Statistical Methods, 2nd Edition. Wiley-Interscience, 1999.

[145] Y. Hochberg and A. C. Tamhane. Multiple comparison procedures. John Wiley & Sons, Inc., New York, NY, USA, 1987.

[146] G.P. Henze, G.S. Pavlak, A.R. Florita, R.H. Dodier, and A.I. Hirsch. An energy signal tool for decision support in building energy systems. Applied energy, 138:51–70, 2015.

[147] L. Wang, S. Greenberg, J. Fiegel, A. Rubalcava, S. Earni, X. Pang, R. Yin, S. Woodworth, and J. Hernandez-Maldonado. Monitoring-based hvac commissioning of an existing office building for energy efficiency. Applied Energy, 102:1382–1390, 2 2013.

[148] Y. Zhao, S. Wang, and F. Xiao. Pattern recognition-based chillers fault detection method using support vector data description. Applied Energy, 112:1041–1048, 12 2013.

[149] R. Gulbinas, R. K. Jain, and J. E. Taylor. Bizwatts: A modular socio-technical energy management system for empowering commercial building occupants to conserve energy. Applied Energy, In-Press, 2014.

[150] S. Kiluk. Dynamic classification system in large-scale supervision of energy efficiency in buildings. Applied Energy, 132:1–14, 11 2014.

[151] K. Bauermann, S. Spiecker, and C. Weber. Individual decisions and system development integrating modelling approaches for the heating market. Applied Energy, 116:149–158, 3 2014.

[152] G. Pavlak, A. Florita, G. Henze, and B. Rajagopalan. Comparison of traditional and bayesian calibration techniques for gray-box modeling. Journal of Architectural Engineering, 20(2):04013011, 2014.

[153] S. De Wit and G. Augenbroe. Analysis of uncertainty in building design evaluations and its implications. Energy and Buildings, 34(9):951 – 958, 2002.

[154] C. Diakaki, E. Grigoroudis, N. Kabelis, D. Kolokotsa, K. Kalaitzakis, and G. Stavrakakis. A multi-objective decision model for the improvement of energy efficiency in buildings. Energy, 35(12):5483 – 5496, 2010.

[155] C. Diakaki, E. Grigoroudis, and D. Kolokotsa. Performance study of a multi-objective mathematical programming modelling approach for energy decision-making in buildings. Energy, 59:534 – 542, 2013.

[156] M.Y. Rafiq. Importance of pareto optimum solutions in making informed decisions in engineering design. In Eighth International Conference on Computing in Civil and Building Engineering, volume 2, pages 1325 – 1333, Stanford, CA, USA, 2000. ASCE.

[157] J.Douglas Balcomb and Adrianne Curtner. Multi-criteria decision-making process for buildings. In Energy Conversion Engineering Conference and Exhibit, volume 1, pages 528 – 535, Las Vegas, NV, USA, 2000. IEEE.

[158] O.O. Ugwu, C.J. Anumba, L. Newnham, and A. Thorpe. Agent-oriented collaborative design of industrial buildings. In Eighth International Conference on Computing in Civil and Building Engineering, volume 1, pages 333 – 340, Stanford, CA, USA, 2000. ASCE.

[159] Y. Sun, Y. Heo, M. Tan, H. Xie, C.F. Jeff Wu, and G. Augenbroe. Uncertainty quantification of microclimate variables in building energy models. Journal of Building Performance Simulation, 7(1):17 – 32, 2014.

[160] Z. Li, Y. Heo, and G. Augenbroe. Hvac design informed by organizational simulation. In Eleventh International IBPSA Conference, pages 2198 – 2203, Glasgow, UK, 2009.

[161] A. Takizawa, H. Kawamura, and A. Tani. Simulation of urban system "as it could be" by multiagent model. In Eighth International Conference on Computing in Civil and Building Engineering, volume 2, pages 1566 – 1573, Stanford, CA, USA, 2000. ASCE.

[162] R. Judkoff, D. Wortman, B. O'Doherty, and J. Burch. Methodology for validating building energy analysis simulations. Technical Report NREL Report No. TP-550-42059., National Renewable Energy Laboratory, Golden, CO, 2008.

[163] R. Piotr, B. Laurence, C. Alison, and A.-S. Ahmed. Decision making aid for selection of renewable/sustainable energy systems for buildings. In International Conference on Sustainable Design and Construction, pages 306 – 313, Kansas City, MO, USA, 2012. ASCE.

[164] S. Attia, E. Gratia, A.D. Herde, and J.L.M. Hensen. Simulation-based decision support tool for early stages of zero-energy building design. Energy and Buildings, 49(0):2 – 15, 2012.

[165] J.H. Choi, A. Aziz, and V. Loftness. Decision support for improving occupant environmental satisfaction in office buildings: The relationship between sub-set of ieq satisfaction and overall environmental satisfaction. In Ninth International Conference of Healthy Buildings, page #747, Syracuse, NY, USA, 2009.

[166] W. Zeiler, R. Maaijen, and W. Maassen. Decision support for sustainable, healthier and more productive buildings. In Tenth International Conference Healthy Buildings, pages 1–6, Brisbane, AUS, 2012.

[167] Y. Heo and V.M. Zavala. Gaussian process modeling for measurement and verification of building energy savings. Energy and Buildings, 53:7 – 18, 2012.

[168] M.C. Burkhart, Y. Heo, and V.M. Zavala. Measurement and verification of building systems under uncertain data: A gaussian process modeling approach. Energy and Buildings, 75:189 – 198, 2014.

[169] P. Zhao, S. Suryanarayanan, and M. G. Simoes. An energy management system for building structures using a multi-agent decision-making control methodology. In Industry Applications Society Annual Meeting, pages 1–8, Houston, TX, USA, 2010. IEEE.

[170] Y. Heo, G. Augenbroe, and R. Choudhary. Risk analysis of energy-efficiency projects based on bayesian calibration of building energy models. In Twelth Conference of International Building Performance Simulation Association, pages 2579 – 2586, Sydney, AUS, 2011.

[171] C. De Farias, H. Soares, L. Pirmez, F. Delicato, I. Santos, L.F. Carmo, J. De Souza, A. Zomaya, and M. Dohler. A control and decision system for smart buildings using wireless sensor and actuator networks. European Transactions on Telecommunications, 25(1):120 – 135, 2014.

[172] H.J. Moon and G. Augenbroe. Empowerment of decision-makers in mould remediation. Building Research and Information, 36(5):486 – 498, 2008.

[173] A. Vinh. Computer-based monitoring for decision support systems and disaster preparedness in buildings. Systemics, Cybernetrics and Informatics, 7:51–56, 2009.

[174] A. Filippoupolitis and E. Gelenbe. A decision support system for disaster management in buildings. In Summer Computer Simulation Conference, pages 141 – 147. Society for Modeling & Simulation International, 2009.

[175] G. De Sanctis, K. Fischer, J. Kohler, M. Fontana, and M.H. Faber. A probabilistic framework for generic fire risk assessment and risk-based decision making in buildings. CRC Press, Zurich, Switzerland, 2011.

[176] N. Basso, E. Garavaglia, L. Sgambi, and N. Imagawa. Natural hazards vs. Decision-making processes in buildings life cycle management. CRC Press, New York, NY, USA, 2013.

[177] B.P. Thompson and L.C. Bank. Use of system dynamics as a decision-making tool in building design and operation. Building and Environment, 45(4):1006 – 1015, 2010.

[178] G. Augenbroe, D. Castro, and K. Ramkrishnan. Decision model for energy performance improvements in existing buildings. Journal of Engineering, Design and Technology, 7(1):21 – 36, 2009.

[179] D. Kolokotsa, C. Diakaki, E. Grigoroudis, G. Stavrakakis, and K. Kalaitzakis. Decision support methodologies on the energy efficiency and energy management in buildings. Advances in Building Energy Research, 3(1):121 – 146, 2009.

[180] S. Das, M.Y.L. Chew, and K.L. Poh. Multi-criteria decision analysis in building maintainability using analytical hierarchy process. Construction Management and Economics, 28(10):1043 – 1056, 2010.

[181] P. Gultekin, C.J. Anumba, and R.M. Leicht. Towards an integrated process model and decision support system for high performance green retrofits. In Architectural Engineering Conference 2013, pages 912 – 923, State College, PA, United states, 2013.

[182] H. Mohseni, S. Setunge, G.M. Zhang, and R. Wakefield. Condition monitoring and condition aggregation for optimised decision making in management of buildings. Applied Mechanics and Materials, 438-439:1719 – 1725, 2013.

[183] S. Lee, Y. Liu, S. Chunduri, R.L. Solnosky, J.I. Messner, R.M. Leicht, and C.J. Anumba. Development of a process model to support integrated design for energy efficient buildings. In International Conference on Computing in Civil Engineering, pages 261 – 268, Clearwater Beach, FL, USA, 2012.

[184] S.H. Lee, G. Augenbroe, J.-K. Lee, and F. Zhao. A design methodology for energy infrastructures at the campus scale. Computer-Aided Civil and Infrastructure Engineering, 28(10):753 – 768, 2013.

[185] B.D. Lee, Y. Sun, G. Augenbroe, and C.J.J. Paredis. Towards better prediction of building performance: A workbench to analyze uncertainty in building simulation. In Thirteenth International Conference of the International Building Performance Simulation Association, pages 1231 – 1238, Chambery, France, 2013.

[186] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, and M.J. Witte. Energyplus: creating a new-generation building energy simulation program. Energy and Buildings, 33(4):319–331, 2001.

[187] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, M. Halverson, D. Winiarski, B. Liu, M. Rosenberg, J. Huang, M. Yazdanian, and D. Crawley. Commercial reference building models of the national building stock. Technical report, U.S. Department of Energy, 2010.

[188] J.E. Braun and N. Chaturvedi. An inverse gray-box model for transient building load prediction. HVAC&R Research, 8(1):73–100, 2002.

[189] N. Chaturvedi, J.E. Braun, and R. Bernhard. Analytical tools for dynamic building control: Implementation of thermal storage in building mass. Technical report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2000.

[190] R. Judkoff, J.D. Balcomb, C.E. Hancock, G. Barker, and K. Subbarao. Side-by-side thermal tests of modular offices: A validation study of the stem method. NREL Report TP-550-23940, National Renewable Energy Laboratory, 2000.

[191] K. Subbarao. Pstar - primary and secondary terms-analysis and renormalization: A unified approach to building and energy simulations and short-term testing; a summary. NREL Report TR-254-3347, National Renewable Energy Laboratory, 1988.

[192] K. Subbarao, J.D. Burch, C.E. Hancock, A.B. Lekov, and J.D. Balcomb. Short-term energy monitoring (stem): Application of the pstar method to a residence in fredericksburg, virginia. NREL Report TR-254-3356, National Renewable Energy Laboratory, 1988.

[193] K. Subbarao. Pstar - primary and secondary terms analysis and renormalization: A unified approach to building energy simulations and short-term monitoring. NREL Report TR-254-3175, National Renewable Energy Laboratory, 1988.

[194] K. Subbarao, D. Mort, and J.D. Burch. Short-term measurements for the determination of envelope retrofit performance. NREL Report TP-253-2639, National Renewable Energy Laboratory, 1985.

[195] EN-ISO. EN-ISO 13790: 2007. International Organization for Standardization (EN), 2007.

[196] C. Bracken, B. Rajagopalan, and J. Prairie. A multisite seasonal ensemble streamflow forecasting technique. Water Resources Research, 46(3):1–12, March 2010.

[197] J.E. Seem. Modeling of heat transfer in buildings. PhD thesis, University of Wisconsin-Madison, 1987.

[198] M.J. Brandemuehl, G. Shauna, and A. Inger. Hvac2 toolkit: Algorithms and subroutines for secondary hvac system energy calculations. ASHRAE, Atlanta, GA, 1993.

[199] U.S. Department of Energy. EnergyPlus Engineering Reference. U.S. Department of Energy, 2010.

[200] E.T. Jaynes. Probability theory: the logic of science. Cambridge university press, 2003.

[201] B.A. Cipra. The best of the 20th century: Editors name top 10 algorithms. SIAM news, 33(4):1–2, 2000.

[202] P. Lauret, F. Miranville, H. Boyer, F. Garde, and L. Adelard. Bayesian parameter estimation of convective heat transfer coefficients of a roof-mounted radiant barrier system. Journal of Solar Energy Engineering, 128:213 – 225, May 2006.

[203] A.T. Booth, R. Choudhary, and D.J. Spiegelhalter. Handling uncertainty in housing stock models. Building and Environment, 48(0):35–47, February 2012.

[204] D. Sivia and J. Skilling. Parameter estimation II. In Data Analysis: A Bayesian Tutorial, chapter 3, pages 35 – 77. Oxford University Press, 2006.

[205] W.J. Karplus. The spectrum of mathematical modeling and systems simulation. Mathematics and Computers in Simulation, 19(1):3–10, 1977.

[206] G.L. Bretthorst. Bayesian spectrum analysis and parameter estimation, volume 48. Springer Science & Business Media, 2013.

[207] T.J. Loredo. From laplace to supernova sn 1987a: Bayesian inference in astrophysics. In Maximum entropy and Bayesian methods, pages 81–142. Springer, 1990.

[208] C. Kamath. Understanding wind ramp events through analysis of historical data. In IEEE PES T&D 2010, pages 1–6. IEEE, 2010.

[209] H. Zheng and A. Kusiak. Prediction of wind farm power ramp rates: A data-mining approach. Journal of solar energy engineering, 131(3):031011, 2009.

[210] B.-M. Hodge, M. Hummon, and K. Orwig. Solar ramping distributions over multiple timescales and weather patterns. Technical report, National Renewable Energy Laboratory, 2011.

[211] C.W. Hansen, J.S. Stein, and A. Ellis. Statistical criteria for characterizing irradiance time series. Technical report, Sandia National Laboratories, 2010.

[212] E.H. Bristol. Swinging door trending: Adaptive trend recording? In ISA National Conf. Proc., 1990, pages 749–754, 1990.

Appendix A

A Summary of the Model-Based Benchmarking (ModBen) Project

The ModBen project [16] had the objective of developing practical methods and tools to help identify potential building energy savings in a cost effective, timely, and permanent manner. A top-down methodology was invented to allow the automated analysis of operational building energy performance data. Simple approaches were seen as fundamental to the scalable optimization of energy-efficient building operations, and were verified through monitoring. The four-step procedure for ModBen's continuous quality control system, progressing from course to finely-detailed analytics, can be broken down as follows:

(1) Benchmarking — to establish current operational building energy performance

(2) Data acquisition and storage — utilizing existing systems and specified measurements

(3) Fault detection and optimization — driven by human experts familiar with the building

(4) Operations monitoring — to verify the operational building energy performance

To evaluate the methods and tools, five demonstration buildings were evaluated. In terms of assessing their whole-building energy performance, the core focus was on steps 3–4 of the top-down procedure; however, experts' specification of a clearly defined scope of measurements (step 2) was essential, and savings are calculated relative to current performance levels (step 1). The methods and tools for each building's fault detection and optimization were developed on the basis of a "minimum data set," assumed to span and observe all major energy processes, and tested

accordingly. Tools were integrated during the project with the help of industrial partners, including their products and services.

The following bullet points summarize the ModBen project's key findings with respect to the five demonstration buildings.

- Through the continuous monitoring of operations, energy savings of 5-10% could be achieved in the demonstration buildings with simple methods. The investment in data collection and processing yielded static payback periods ranging from 1–3 years. The simple methods were deemed suitable for buildings whose yearly energy costs are €30,000 or more.

- Despite the relatively sparse nature of the minimum data set, the data acquisition, processing, and transmission system proved to be a major obstacle for the approach — whether a building automation system (BAS) was present or not.

- The continuous acquisition and storage of data, concerning energy consumption and operational parameters, is by no means the state of the art.

- Extending BAS data with additional high time-resolution data streams has proven to be complex and costly. In general, BASs are unprepared for operational analyses. Existing systems are helpful, however, in vetting data and cross-validation.

- Manual fault detection and diagnosis by means of intelligent data visualization has been used extensively and successfully in the initial analysis of a given building's energy performance. However, despite the automatic generation of visualizations, experts' insight and knowledge cannot be understated. A systematic application of the insights into daily operations is unavoidable and an open research question.

- From a practical perspective, black-box models have proven effective for simple and automatic detection of unusual operating conditions.

- With model-based methods for fault detection and diagnosis, followed by optimization, it was apparent white-box or gray-box models offer great potential for operational analyses;

however, in realistic applications, and with regard to model preparation and calibration, they are both still too expensive (i.e. time investment) for practical applications.

- Interoperability is an important aspect for the standardized and seamless electronic exchange of information, although generally not critical for building sector applications.

- Static measurements and metadata, as related to the building's structure, envelope, mechanical equipment, and floor plans, can be immensely helpful to understanding operations.

- Energy performance rating systems or certifications are not widely adopted due to their voluntarily nature, and do not include the data necessary to perform the analyses; they are typically not suitable for deriving target (energy consumption) values either.

- In addition to technical issues, a number of organizational concerns had to be addressed; e.g., privacy, data protection, and general organizational bureaucracy. These impediments have hindered and slowed processes. Within the demonstration buildings, there was clearly an art to process – drawing on the participation of both the organization and staff – beyond operational analyses, pure technical processing of data traffic, and performance assessments. To be successful it was necessary to be thoroughly involved in all processes.

- Adaptation to ever-changing circumstances was crucial, even with short notice and unanticipated circumstances. Investments in energy performance measures for existing buildings must have correspondingly short refinancing periods. Here the focus was the optimization of operations, even under difficult and dynamic usage constraints.

The following future research has been noted as being of interest to extending the results of the ModBen project.

Whole-Building Fault Detection and Diagnosis (FDD): To apply whole-building FDD it may be necessary to focus on typical and common systems, including their subsystems, like heating circuits, ventilation, etc. The minimum dataset was found to be a great start-

ing point for analysis; however, further standardization and representations are required for a wider range of typical systems.

Interoperability: Many operational approaches suffer from the lack of complete, current, and accurate documentation of the building and its systems. Thus, the ability to harness state of the art modeling and simulation tools for model-based approaches is limited. Development of calibrated and validated models is too costly and inaccurate in practical applications. Energy performance rating systems and certifications are helpful on the whole-building level and confirm, or partially confirm, the overall state of the building. Combining these insights with a model-based approach suffers from the availability of comprehensive and conclusive data, as well as uniform standards for labeling measurement data or control points.

Treatment of Modeling Uncertainties: The modeling and simulation of buildings requires a number of inputs that are inherently uncertain. A typical example is occupant behavior; e.g. presence of occupants, loading density, heating and ventilation behavior, etc. In the context of these uncertainties, stochastic optimization approaches are an open research topic for buildings. Such approaches are fundamentally different from classic, deterministic simulations and have the advantage of explicit uncertainty quantification or prediction with ranges. It is envisioned probabilistic approaches will become the norm for forward and inverse building energy modeling tools.

Organizational Aspects: In addition to the technical aspects of integrating operational building energy performance, organizational aspects play an important role in the success of the operational analysis of buildings. Responsibilities need to be clarified, with necessary processes documented and tracked. There seems to be a need for unification with other business processes, including the development of appropriate tools and tracking processes.

Appendix B

Fundamentals of Bayesian Inference

In this appendix, the necessary background on Bayesian inference is provided as review; its development summarizes Bayesian-focused research as expounded by Jaynes [200], Bretthorst [206], and Laredo [207].

B.1 Derivation of Bayes' Theorem

With the development of probability theory being attributed to Bernoulli and Laplace, Bayes' theorem can be derived through the familiar axioms. The sum rule states that the probability an even is true or false must sum to unity. That is, assuming an even is known to exist in space 'Z', what is the probability that it also exists in space 'X' or (mutually exclusive) space 'Y' when it is known that 'X' and 'Y' are subspaces of space 'Z'? The sum rule can be stated as:

$$p(X|Z) + p(\bar{X}|Z) = 1, \tag{B.1}$$

where the horizontal bar represents the negation of the proposition, and the vertical bar a conditional statement.

The product rule is also helpful. It expresses the relationship between spaces as various subspaces of each other. For instance, what is the probability that an event exists in 'X' and (mutually inclusive) 'Y' if one knows the event is a subspace of 'Z'?

$$p(XY|Z) = p(X|YZ)p(Y|Z) \tag{B.2}$$

All relationships between probabilities can be derived using these two equations. For instance, the probability that an event exists in 'X' or 'Y', if one knows the event is a subspace of 'Z', can be calculated as:

$$p(X+Y|Z) = p(X|Z) + p(Y|Z) - p(XY|Z). \tag{B.3}$$

It should be kept in mind that the arguments for a probability symbol are propositions and not numbers. Furthermore, the operations inside the parentheses are logical operations: the logical 'and' is the multiplication of probabilities and is represented by placing propositions side-by-side, and the logical 'or' is the sum of probabilities and represented by the '+' symbol.

To derive Bayes' theorem we recognize that XY and YX are equal, and Equation B.2 can be rewritten as $p(X|YZ)p(Y|Z) = p(Y|XZ)p(X|Z)$ and solving for $p(X|YZ)$ we get:

$$p(X|YZ) = p(X|Z)\frac{p(Y|XZ)}{p(Y|Z)} \tag{B.4}$$

This is Bayes' theorem.

The theorem is used for hypothesis testing. To see how this can be done we let $X = H$, a hypothesis we want to test, $Y = D$, some data that is relevant to hypothesis testing, and $Z = K$, some knowledge-base that indicates the way in which H and D are related. Bayes' theorem can now be written as:

$$p(H|DK) = p(H|K)\frac{p(D|HK)}{p(D|K)} \tag{B.5}$$

This allows the plausibility assessments regarding states of knowledge to be adjusted with the acquisition of data, and represents a form of learning. The posterior probability of H is obtained by multiplying the prior probability $p(H|K)$ by the probability of data assuming a true hypothesis, $p(D|HK)$, and dividing it by the probability that the data would have been observed nonetheless, $p(D|K)$. This is just the state of knowledge-base before and after the observation of data.

The knowledge-base 'K' is the space where the physics of interest will be explored. Of all the physical phenomena, we are interested in exploring a small space of unknown physical relationships – our hypothesis space 'H'. The data 'D' either reinforces or negates the hypotheses that have been formed through Bayes' theorem.

B.2 Bayesian Parameter Uncertainty Quantification

Many hypotheses are possible, and inferences can be made about a parameterized model. The parameters can be discrete or continuous, but as an example we will take Θ as a single parameter for descriptive purposes. As an uncertainty quantification problem, a model is assumed true for some unknown parameter value with constraints imposed as necessary. The hypothesis space is therefore all the possible values that the parameter may take. The data is in a form that allows inferences to be made and the probability, in distribution form, of any given Θ.

Bayes' theorem can be used to address the uncertainty quantification problem by calculating the probabilities directly. If 'D' represents a proposition asserting the values of data observed, and 'H' the proposition that $\Theta = \Theta'$ asserting that a possible parameter value is the true value, the probabilities can be calculated. It should be noted that 'K' represents the knowledge-base necessary to convey all the necessary relationships (e.g., physics), constraints, and prior information. Bayes' theorem can therefore be written:

$$p(\Theta|DK) = p(\Theta|K)\frac{p(D|\Theta K)}{p(D|K)} \tag{B.6}$$

In order to use the equation, the three probabilities on the right-hand side must be specified. The prior $p(\Theta|K)$ and the likelihood $p(D|\Theta K)$ must be assigned a priori. That is, the prior represents already known information about the problem that can be specified in the model in parameter form. The likelihood is calculated using the measured data and expectation according to the model. The model must be true for some value of its parameter(s). Therefore, one proposition of $\Theta = \Theta 1'$ or $\Theta = \Theta 2'$ or $\Theta = \Theta 3'$, etc. must be true. This is written as:

$$\begin{aligned} p(D[\Theta_1 + \Theta_2 + \ldots]|K) &= p(D|K)p(\Theta_1 + \Theta_2 + \ldots|K) \\ &= p(D|K) \end{aligned} \tag{B.7}$$

Expanding the left-hand side, the following is true:

$$\begin{aligned} p(D[\Theta_1 + \Theta_2 + \dots]|K) &= p(D\Theta_1|K) + p(D\Theta_2|K) + \dots \\ &= \sum_i p(D\Theta_i|K) \\ &= \sum_i p(\Theta_i|K)p(D|\Theta_i K). \end{aligned} \tag{B.8}$$

Equations B.6 and B.7 can be combined in the following form:

$$p(D|K) = \sum_i p(\Theta_i|K)p(D|\Theta_i K). \tag{B.9}$$

Equations B.8 express $p(D|K)$ as we desire, and is expressed in terms of a prior and likelihood. In the uncertainty quantification problem, the denominator is just a normalization constant for the posterior. It expresses the all possible scenarios for the parameter. The calculation of the normalization constant is what makes Bayesian inference difficult to implement in practice, as the simulation of millions of building energy models are necessary to ensure that the normalization constant is calculated fully in that it converges. Monte Carlo approaches are necessary to fully explore the parameter space. The results are summarized by considering the marginalization of the parameter over all possible parameter values. When dealing with continuous parameters, the summation becomes an integral and can be written:

$$p(D|K) = \int p(\Theta|K)p(D|\Theta K)d\Theta. \tag{B.10}$$

To summarize the inferences, the posterior can be visualized graphically or in table form. This is complicated by the fact the parameter space can be high dimensional. When dealing with numerous parameters, we may be interested in only a subset of these parameters. To average over parameters:

$$p(\Theta|DK) = \frac{1}{p(D|K)} \int p(\beta|K)p(\Theta|\beta K)p(D|\Theta\beta K)d\beta. \tag{B.11}$$

We marginalize over nuisance parameters. That is, if a problem has two parameters: Θ and β, but we are only interested in Θ. This of great practical use because the dimensionality can be reduced and posterior distributions can be plot two-dimensionally for any given Θ. Marginalization

is the advantage of the Bayesian perspective because it allows the summarize of simulation results and inference in a compact form.

B.3 Bayesian Model Comparison

The Bayesian methods presented thus far assume that the model used for parameter inference is the correct model. However, it may be the case that numerous models can approximate the physics of any given system and it is of interest to assess which model is the superior model for prediction. The results follow Ockham's razor: simpler models are preferred unless more complicated models provide significantly better fit to the data.

To use Bayes' theorem for model comparison, the knowledge-base asserts that one of the set of models must be the "true" model. For m models (m=1 to M), Bayes' theorem can be used to calculate the probability a model is the true model:

$$p(m|DK) = p(m|K)\frac{p(D|mK)}{p(D|K)}, \tag{B.12}$$

where, if no model is favored over the others a uniform prior distribution is specified.

To calculate $p(D|mK)$, the marginal likelihood for model m must be calculated. The parameters of model m are known beforehand:

$$p(D|mK) = \int p(\Theta_m|K_m)p(D|\Theta_m K_m)d\Theta_m. \tag{B.13}$$

To calculate $p(D|K)$, a marginalization over the m models must be calculated:

$$p(D|K) = \sum_m p(m|K)p(D|mK). \tag{B.14}$$

Just as will the uncertainty quantification problem, $p(D|K)$ is simply the normalization constant. However, in model comparison problems its calculation can be avoided by considering the ratio of probabilities of the models, rather than directly the probabilities themselves. This ratio is called the odds, and the odds for favor of model m_1 over model m_2 can be calculated from:

$$\begin{aligned} O_{m_1,m_2} &= \left[\frac{p(m_1|K)}{p(m_2|K)}\right] \frac{\int p(\Theta_{m_1}|K_{m_1})p(D|\Theta_{m_1}K_{m_1})d\Theta_{m_1}}{\int p(\Theta_{m_2}|K_{m_2})p(D|\Theta_{m_2}K_{m_2})d\Theta_{m_2}} \\ &= \left[\frac{p(m_1|K)}{p(m_2|K)}\right] B_{m_1,m_2}, \end{aligned} \tag{B.15}$$

where the factor in brackets is called the prior odds. The B_{m_1,m_2} is called the Bayes' Factor, which is just the ratio of prior predictive probabilities.

The approach presented is different to the model parameter uncertainty quantification, but is necessary for not only determining what the most probable parameters are but also the most probably model. Using both of these tools allows the most accurate representation of any building energy process through the use of data.

B.4 Requirements for Computation

B.4.1 Integral Estimation via Monte Carlo Method

With a range of faults that can be observed in building operation and in recorded data, a practical method for statistical inference is required for the concerns of whole-building fault detection and diagnosis. The construction of probability models for observed quantities, and for quantities that inferences are desired, are quite helpful in detecting whole-building faults and pointing toward the most probable diagnosis. Bayesian methods explicitly use probability for quantifying uncertainty, and therefore are useful for the present concerns.

The Bayesian data analysis process is divided into three general steps:

(1) Set up the full probability model for all observable and unobservable quantities in the problem.

(2) Calculating and interpreting the posterior distribution – the conditional probability distribution of the unobserved quantities given the observed data.

(3) Evaluating the fit of model and the implications of the posterior distribution.

A major issue in Bayesian analysis is the required conditioning on known data y, as part of Bayes' rule, to yield the posterior density so that inferences can be made. This requires evaluating complex integrals.

Monte Carlo simulation is a stochastic technique to evaluating complex problems. That is, random numbers are used to build probability models that are used to investigate problems which, due to their nature, can have multiple "solutions" due to uncertainties in variables/parameters that are beyond the control (or prior knowledge) of the investigator. The goal is to determine likely scenarios, which can be based on measured data for model fitting (e.g. building model calibration) or could be based on external disturbances that affect the likely bounds of the "system" output that influences the decision making process (e.g. investment risk analysis).

The mean value of a given stochastic variable can be expressed as the integral of the product of a variable and its probability density function. The process can be viewed in reverse, such that random numbers are generated, evaluated, and estimations of the mean can be made by simple averaging. From either viewpoint, the theory is to evaluate an integral of the form:

$$G = \int_D h(x)f(x)dx \tag{B.16}$$

where the domain, D, with coordinates x is too large and complex for analytic solution, and $f(x)$ is a non-negative function that satisfies:

$$\int_D f(x)dx = 1. \tag{B.17}$$

Thus, f can interpreted as a probability density function, and Equation B.16 as the expectation of $h(x)$, where x is the random variable of interest.

For low-dimensional spaces, the Monte Carlo method is inefficient, and there are better (non-random) numerical approximations for computing the integral of Equation B.16. The typical methods is to discretize the space D by x_k for $k = 1, 2, \ldots, n$, and approximate the area under the curve. For an M dimensional space in domain D, the error in the approximation decreases with N as $N^{-1/M}$, and is too slow for practical use in complex problems. However, it can be proved that the error in the Monte Carlo method decreases as $N^{-1/2}$ regardless of the dimension of D. This is the main advantage of the Monte Carlo method.

Markov Chains are helpful for proving that the Monte Carlo simulation converges to the target distribution, and that is why they are helpful. Proof to come.

B.4.2 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) simulation is a general means toward developing the probability model. The probability model building process is iterative: θ values are drawn from an approximate distribution and corrections are made to better approximate the target posterior distribution, $p(\theta|y)$. The sequential sampling depends only on the previous draw, and thus the draws form a Markov chain. The Markov property is needed to prove that the approximate distribution converges to a stationary distribution that is equal to the target distribution. Each sequence starts at a chosen θ^0, and for each t a new θ^t is determined from a transition distribution $T_t(\theta|\theta^{t-1})$ that only depends on the previous draw.

The following subsections briefly discuss three major MCMC methods.

B.4.3 Metropolis Algorithm

The first method developed for acquiring a sequence of random samples from a probability distribution was the Metropolis algorithm. A variation of a random walk, it uses an acceptance/rejection rule until the approximate distribution converges to the target distribution.

Pseudocode for the Metropolis algorithm is as follows:

- From a starting distribution $p_0(\theta)$ use a starting point θ^0, such that $p(\theta^0|y) > 0$.
- $t = 1, 2, \ldots$
  - Sample a proposal θ^* from a *jumping (proposal) distribution* $J_t(\theta^*|\theta^{t-1})$ at time t.
  - Calculate the ratio of densities,

$$r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)}. \tag{B.18}$$

  - Set

$$\theta^t = \begin{cases} \theta^* & with\ Pr(min(r,1)) \\ \theta^{t-1} & otherwise. \end{cases} \tag{B.19}$$

The Metropolis algorithm requires that the jumping (proposal) distribution is symmetric. That is, $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$ for all θ_a, θ_b, and t. The jumping distribution seeks to find a point at random in a specified area close to the previous θ; This is directly related to Simulated Annealing in Optimization Theory. An issue with the Metropolis algorithm is finding a suitable jumping distribution so that the random search is effective. Otherwise, if there is poor scaling, the algorithm could take a very, very long time to converge to the target distribution. Standard practice is to "tune" the variance of the jumping function until there is a proper acceptance rate for the algorithm.

B.4.4 Metropolis-Hastings Algorithm

The Metropolis-Hasting algorithm generalizes the Metropolis algorithm so that the jumping rule J_t is not required to be symmetric. The asymmetric jumping rule is possible by replacing the ratio r in Equation B.18 of the Metropolis algorithm with a ratio of ratios:

$$r = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)}. \tag{B.20}$$

Allowing for an asymmetric jumping rule can increase the speed of the random walk.

B.4.5 Gibbs Sampler

Gibbs sampling, also known as *alternating conditional sampling*, can be thought of as a special case of the Metropolis-Hasting algorithm. It is particularly useful in multidimensional problems, and achieves this via conditional sampling of the variable/parameter vector that has been divided into k components or subvectors, so that $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. At each iteration, the Gibbs sampler cycles through the components/subvectors of θ and draws conditional on the other values; this requires k steps for each iteration t. An ordering of the k components/subvectors is done at each iteration t, and each θ_j^t is conditionally sampled given the other components:

$$p(\theta_j|\theta_{\neg j}^{t-1}, y), \tag{B.21}$$

where $\theta_{\neg j}^{t-1}$ is all the components of θ, not including θ_j, at their current value:

$$\theta_{\neg j}^{t-1} = (\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^t, \ldots, \theta_k^{t-1}). \tag{B.22}$$

The components/subvectors θ_j are therefore conditionally updated according to the latest values of the other components/subvectors.

Appendix C

Identifying Wind and Solar Ramping Events

Wind and solar power are playing an increasing role in the electrical grid, but their inherent power variability can augment uncertainties in the operation of power systems. One solution to help mitigate the impacts and provide more flexibility is enhanced wind and solar power forecasting; however, its relative utility is also uncertain. Within the variability of solar and wind power, repercussions from large ramping events are of primary concern. At the same time, there is no clear definition of what constitutes a ramping event, with various criteria used in different operational areas. Here, the swinging door algorithm, originally used for data compression in trend logging, is applied to identify variable generation ramping events from historic operational data. The identification of ramps in a simple and automated fashion is a critical task that feeds into a larger work of 1) defining novel metrics for wind and solar power forecasting that attempt to capture the true impact of forecast errors on system operations and economics, and 2) informing various power system models in a data-driven manner for superior exploratory simulation research. Both allow inference on sensitivities and meaningful correlations, as well as quantify the value of probabilistic approaches for future use in practice.[1]

C.1 Introduction

The increasing amounts of wind and solar power capacity being installed in the electrical system are causing more concern from system operators about the variable and uncertain nature of

[1] Appendix C largely derives from Florita et al. [20] and authorship is completely attributed to Anthony R. Florita; the other authors contributed editing.

these generators. To an extent, power system operations are already able to handle variability and uncertainty, e.g., power demand. Existing techniques include regulation reserves, load-following reserves, and sub-hourly economic dispatch. However, in simplistic terms, the uncertainty in load, now coupled with increasing levels of uncertainty in generation, can lead to wider distributions of uncertainty for all variables and parameters of interest; responding to variability under increased uncertainty is all the more difficult. Enhanced wind and solar power forecasting can help address some of these concerns through the reduction of uncertainty faced by the system of interest. Because there are mechanisms in place to handle small amounts of uncertainty and variability, power system operators place primary emphasis on better understanding the impact of extreme events (e.g., large ramps), which can have significant influence on system economics and reliability. Secondary concern is for uniform power forecasting improvements for enhanced planning applications.

Wind and solar ramps can occur at different timescales, geographic scales, and in both the positive and negative directions. Variable generation forecasting can help remove some of the uncertainty involved with the power supply, but may have trouble forecasting large ramping events. The numerical weather prediction models often used for forecasting are generally good at predicting roughly when a ramping event may occur; however, there are two main ways in which inaccurate forecasting of ramp events can lead to large errors: ramp magnitude and timing errors. In ramp magnitude errors, a ramp is forecast, but the actual value changes significantly more/less than was forecast. In ramp timing errors, the actual ramp in power significantly leads/lags the forecast time. Of course, both errors can occur simultaneously, which indicates a poor forecast. It is the hope that offline ramp analyses, coupled with extensive unit commitment and dispatch simulation studies, will allow the synthesis of knowledge for enhanced dispatch in cases of large variable generation power ramps.

The automated identification of ramping events must be computationally inexpensive to justify online applications, but can also help facilitate the improvement of forecasting algorithms by providing metrics on how well ramping events are captured. Kamath analyzed wind ramping events in the Bonneville Power Administration area using two definitions of ramping [208]. The

first definition was simply the slope of change between two points; the second considered the minimum and maximum values of generation between two points. Zheng and Kusiak [209] focused on forecasting wind power ramping events. They employed the rate of change of wind plant power over a 10-minute interval to define ramps. Hodge et al. [210] used similar fixed-point definitions to identify and characterize the number of ramping events that occurred for solar power at different timescales. Hansen et al. [211] used the swinging door algorithm to characterize irradiance time series data in the Southwest United States. Because of the flexibility and simplicity of the algorithm, both wind and solar power ramps over varying time frames can be identified.

C.2 Swinging Door Algorithm

In this work, we propose the application of an algorithm from the area of data compression, known as the swinging door algorithm [212], to identify wind and solar power ramping events. Its computational and structural simplicity, requiring only one parameter in its definition, are favorable attributes considering its robustness in the face of noisy data.

Ramps are typically extracted through a linear piecewise approximation to the original time series of data. If extracting ramps from measured data, the approximation can be thought of as a disregard for the noise inherent to the measurement process and/or insignificant changes. If extracting ramps from simulated data, the approximation can be thought of solely as disregard for the insignificant changes. In either case, the focus of ramp extraction is placed on the significant linear ramps (in terms of magnitude and duration) present in the dataset.

Mathematically speaking, a ramp is quantified by its instantaneous rate of change, its derivative, $\frac{dG}{dt}$, and is approximated initially by a local ratio of differences: $\frac{dG}{dt} = \frac{G(k)-G(k-1)}{k-(k-1)}$. The discrete-time nature of either the measured or simulated data easily allows such a calculation. However, the point of ramp extraction is to determine a trend in a sequence of local derivatives and the magnitude and duration of such a trend. For example, when considering a time series of power, the local derivative (ramp) from the first two points may be $\frac{3.0}{2-1} = 3MW$ and from the second and third points $\frac{3.2}{3-2} = 3.2MW$. The trend is apparent and the average ramp is 3.1 MW over the three

discrete-time samples. The question of interest is when a particular ramp has started and/or when the local derivative has changed to the point it can no longer be considered part of a particular ramp.

Figure C.1 illustrates a simplified example of a signal and the ramps that may be extracted. Of course, a realistic time series of wind or solar power is much more complicated, but the same strategies and goals for the extraction of ramps apply as described here. The measurement points are discrete-time samples, and the spline fit is included in the figure to show what the continuous process may resemble. The identified ramps are nearly of equal magnitude, but in general this will not be the case. It is somewhat easy to visually discern the ramps (trends) even though it is apparent the sign of the slope can change within a particular ramp. Although noise is inherent to any real measured data, here there are no assumptions about the probability density of a realization and the piecewise linear approximation to the time series is anchored to dominant points. Considering a threshold for the ramp trend and anchoring the piecewise linear approximation to measurement points allows for reduced sensitivity to inflection points and other insignificant fluctuations.

The swinging door algorithm allows the extraction of ramps in a signal, in a piecewise linear fashion, while allowing for consideration of a threshold parameter influencing its sensitivity to ramp variations. The only tunable parameter is ϵ, the width of one door in the algorithm (as shown in Figure C.2) that directly allows the (threshold) sensitivity to noise and/or insignificant fluctuations to be specified. If the tolerance is very low (a small ϵ value), the ramp extraction algorithm will identify many small ramps as it basically traces the original signal, violates the threshold, and starts over. If the tolerance is very high (a large ϵ value), the algorithm will identify a few large ramps as it is under constrained and a large fluctuation is required for the threshold to be violated. In the figure, it should be noted the scale is arbitrary for the purposes of explanation, and in general the signal magnitude is much larger than the scale of the threshold bounds.

From Figure C.2, the swinging door algorithm is briefly described: 1) the initial (dominant) point, or new (ramp segment) iteration of the algorithm, is on the y-axis and threshold doors of width ϵ are placed above and below it; 2) a new point A is acquired and the doors "swing open,"

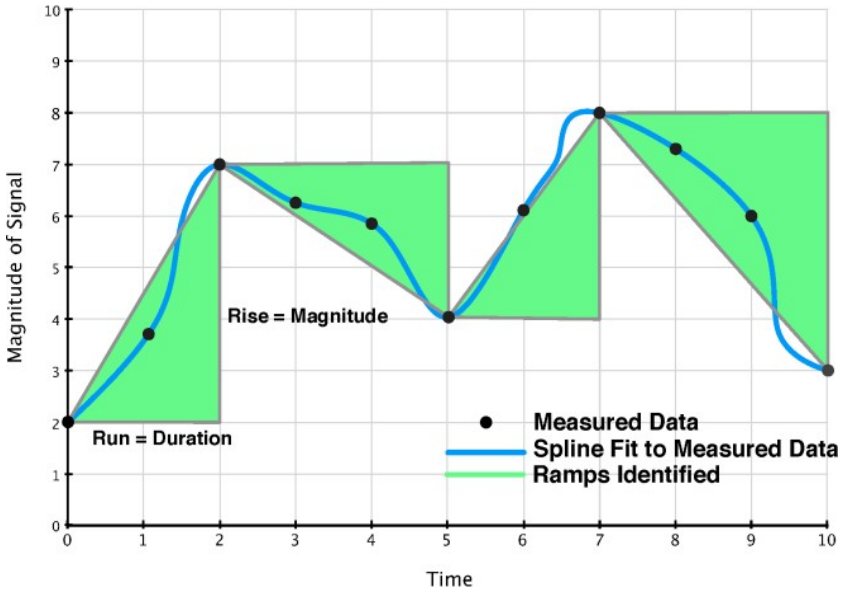

Figure C.1: Example of the piecewise linear approximation to a time series for ramp extraction and analysis; the scale is arbitrary for explanation purposes.

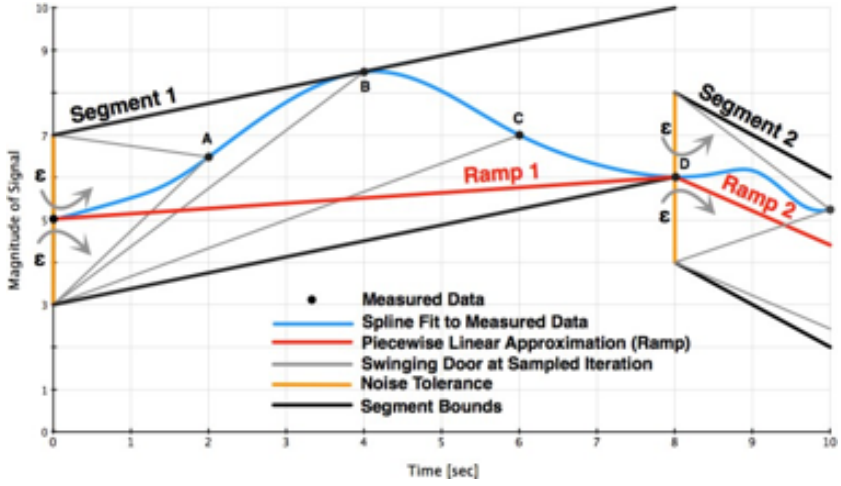

Figure C.2: The swinging door algorithm for the extraction of ramps in power from the time series; the scale is arbitrary for explanation purposes.

as indicated, to include the point i.e., lines are drawn from the doors' hinges to the point; 3) a new point B is acquired and lines are again drawn (updated) to intersect at B; 4) a new point C is acquired, but there has been an inflection in the signal, and the swinging doors open only to accommodate new points in a ramp segment iteration, so the top door (extended line) remains in its angle position above C and the lower door line is drawn to point C further extension of the lines would result in an intersection at some point in the future; 5) a new point D is acquired; again the top door (line) angle position is not updated, and the lower door line is drawn to point D. The lines are now parallel (or do not intersect in the future), which starts a new iteration of the algorithm i.e. the threshold has been exceeded when the line angle from the hinges to their most open position is greater than or equal to parallel. The threshold could be violated somewhere between C and D, but because of the discrete-time nature of the approximation, a new iteration would still start at D. The piecewise linear approximation (the ramp, shown in red) starts at the end of the previous iteration (dominant point) and ends when the threshold is exceeded (next dominant point).

Ramp sign changes are an indicator of fluctuation, but it is not obvious what an insignificant fluctuation is when considering noisy measured data and/or actual (but slight) power variations. There are two applications that are noted for defining the threshold and thus what is considered an insignificant fluctuation: 1) according to the accuracy of the measurement device as defined by its distribution of measurement uncertainty, or 2) according to the utility of the measure as defined by power system economics and its relative importance in driving operations. In this work, neither application is explicitly employed, but the ϵ value varied to explore the sensitivity of ramp events extracted according to its value. Specifically, the ϵ value is set to a percentage of the maximum capacity observed in the time frame of interest.

Figure C.3 shows a typical example in the extraction of ramps from a large wind farm over a two-day period. The power profile, composed of hourly data, is variable but somewhat smooth because of the diversity in power from individual turbines aggregating to cancel high-frequency variability, combined with time-averaged power output over the hour. Therefore, a rather high tolerance, ϵ value of 10% of maximum capacity, was used and provided an accurate piecewise linear

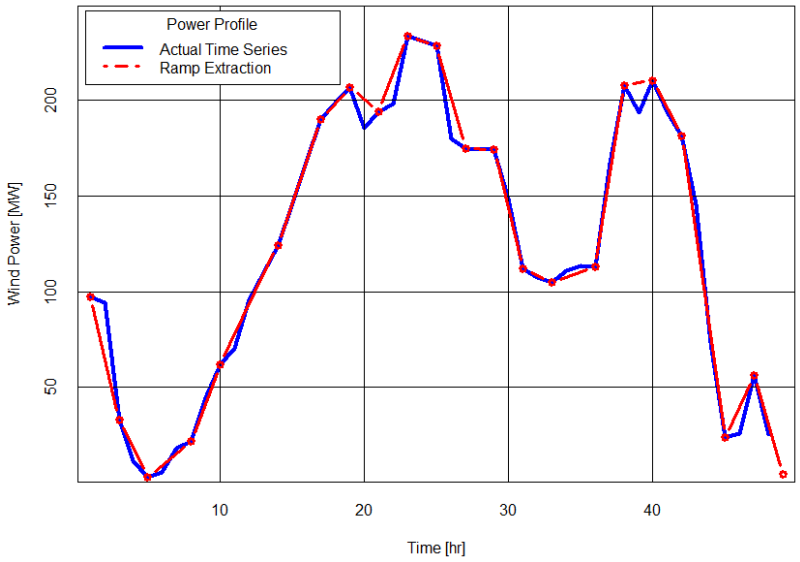

Figure C.3: Typical example of ramp extraction from two days of power at a large wind farm, showing up and down ramps of large, medium, and insignificant nature.

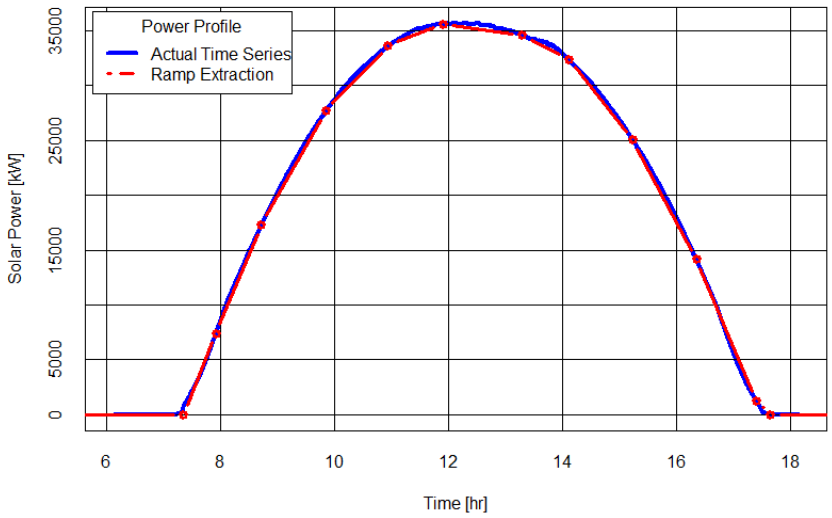

Figure C.4: Typical example of ramp extraction from the first of two days of power at a PV solar plant, showing a clear day leading to a smooth profile.

approximation to the wind power profile.

Figures C.4 and C.5 show typical examples in the extraction of ramps from a solar plant over a two-day period, both using data sampled on a one-minute basis; first, the clear day of Figure C.4, followed by the somewhat cloudy day of Figure C.5. The power profile is smooth in Figure C.4 and shows high-frequency variability in Figure C.5. A rather low tolerance, ϵ value of 1% of maximum capacity, was used and provided an accurate piecewise linear approximation to the solar power profile. However, it is apparent the ϵ choice introduces tradeoffs between the count of ramps and their approximation accuracy. That is, a clear day may be adequately described by fewer piecewise segments, whereas a cloudy day may require more for an adequate description. Economics of the system under consideration will likely determine the choice of ϵ.

C.3 Wind and Solar Data

To showcase the use of the swinging door algorithm for wind and solar power ramp detection, it was applied to various datasets. Wind data came from a wind plant in the Xcel Colorado territory with an approximate capacity of 300 MW; the discrete-time sample was 1 minute. The solar data

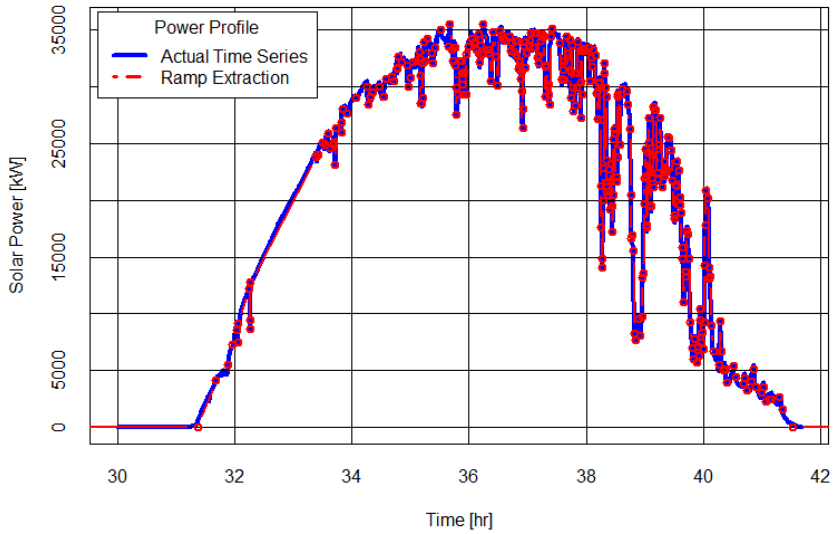

Figure C.5: Typical example of ramp extraction from the second of two days of power at a PV solar plant, showing up and down ramps because of clouds.

came from Oahu and Maui, Hawaii, in association with the Hawaiian Solar Integration Study; the discrete-time sample was 1-second.

C.4 Results

The resolution of the extracted ramp events is a function of ϵ, which is informed by application specifics. The utility of a given magnitude of ramp event (as part of power system economics) was not considered here, but it is the subject of ongoing research toward understanding the probabilistic relationships of various systems. Typical wind and solar power examples were provided in this section, but the time resolution, geographic diversity, and extent of smoothing from plant aggregation were limited to the data available.

Ramp extractions were visualized by rise-run distributions. Figures C.6 and C.7 give the bivariate distributions of wind and solar power, respectively, as a function of various ϵ values; in both figures, the ϵ value was, from top-to-bottom subplots, equal to 1, 2, 3, and 5% of the maximum capacity observed in the month of December.

From the wind power ramp extraction of Figure C.6 (i.e., rise [MW] versus run [min]), it is noted that with lower tolerance, more ramps of longer duration were extracted. This would be expected, but it is also interesting to note how the distribution spreads within the more immediate (quick) ramp region. In the solar power ramp extraction of Figure C.7 (rise [kW] versus run [s]), the same trends as the wind example were noted; however, there appeared to be a correlated "what goes up, must come down" pattern to the ramps because of the diurnal nature of solar irradiance. That is, there was an approximate balance of up and down ramps of similar magnitude and duration. Furthermore, the dispersion of ramps was driven by the plant (area) size and the December cloud cover.

As might be expected, smoothing from aggregation was observed in both power datasets and varied according to the size (area) of the total plant. In wind power, the downstream turbines generally experienced slower and more turbulent wind, and spatial correlations in power variability diminished with distance. In solar power, cloud cover seemed to have only influenced a portion of

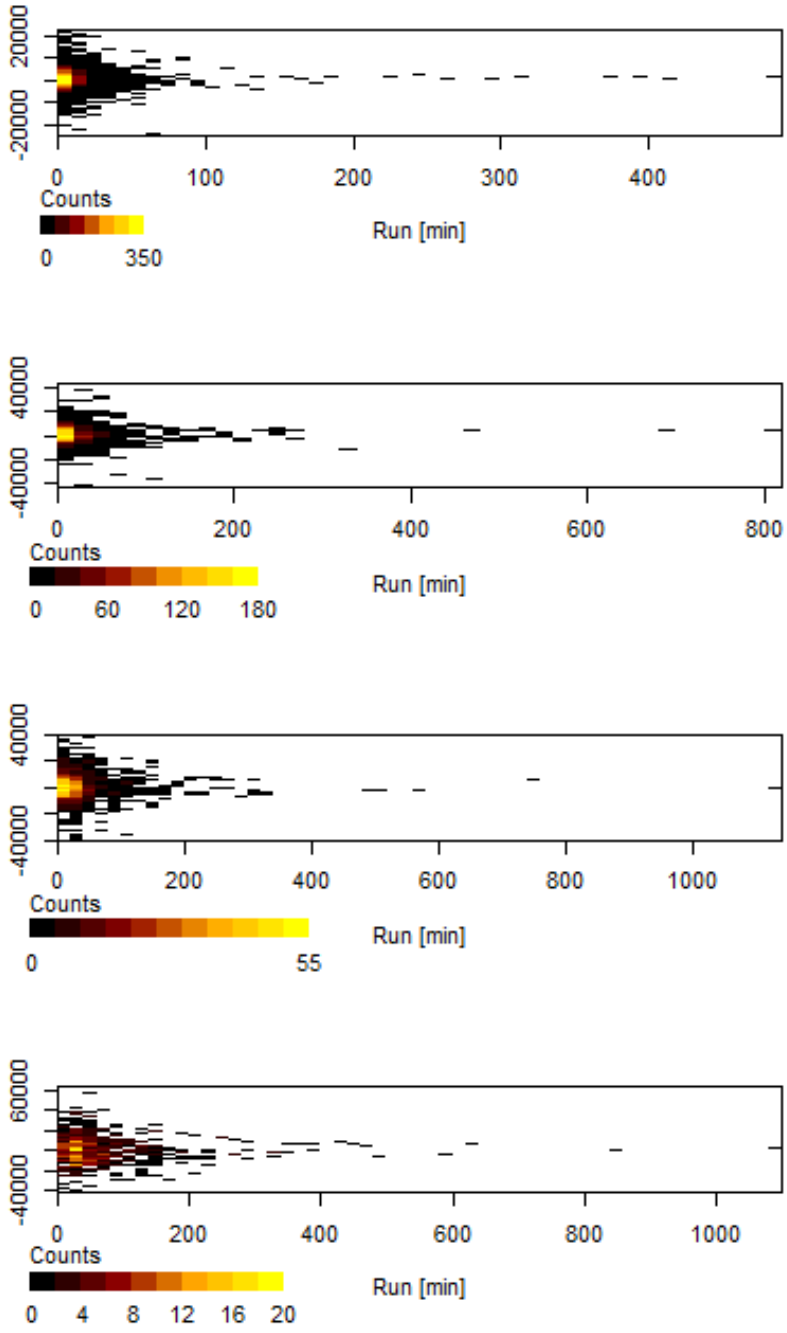

Figure C.6: Bivariate distribution of wind power, ramp rise versus run, as a function of the ϵ value; top subplot is $\epsilon = 1\%$ maximum capacity in December, followed by $\epsilon = 2$, 3, and 5%, respectively.

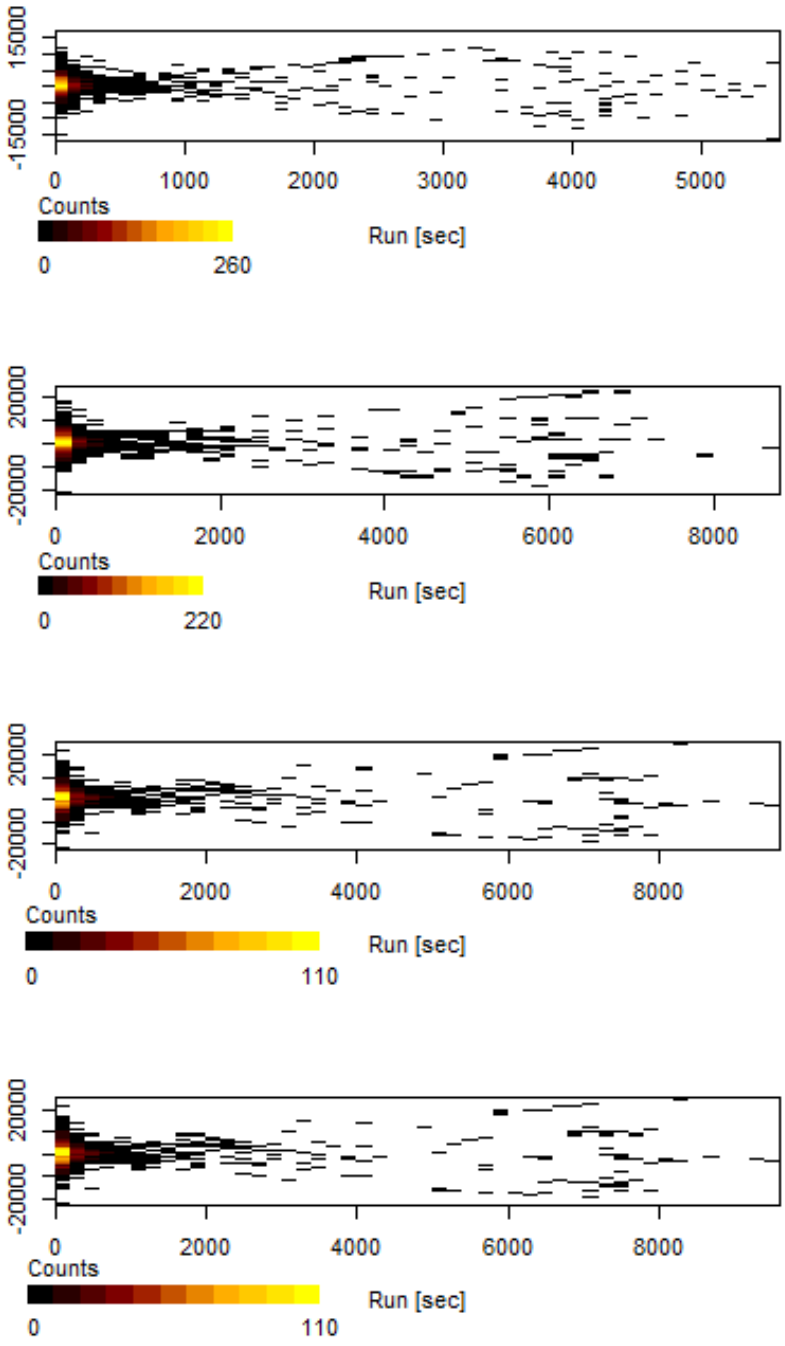

Figure C.7: Bivariate distribution of solar power, ramp rise versus run, as a function of the ϵ value; top subplot is $\epsilon = 1\%$ maximum capacity in December, followed by $\epsilon = 2$, 3, and 5%, respectively.

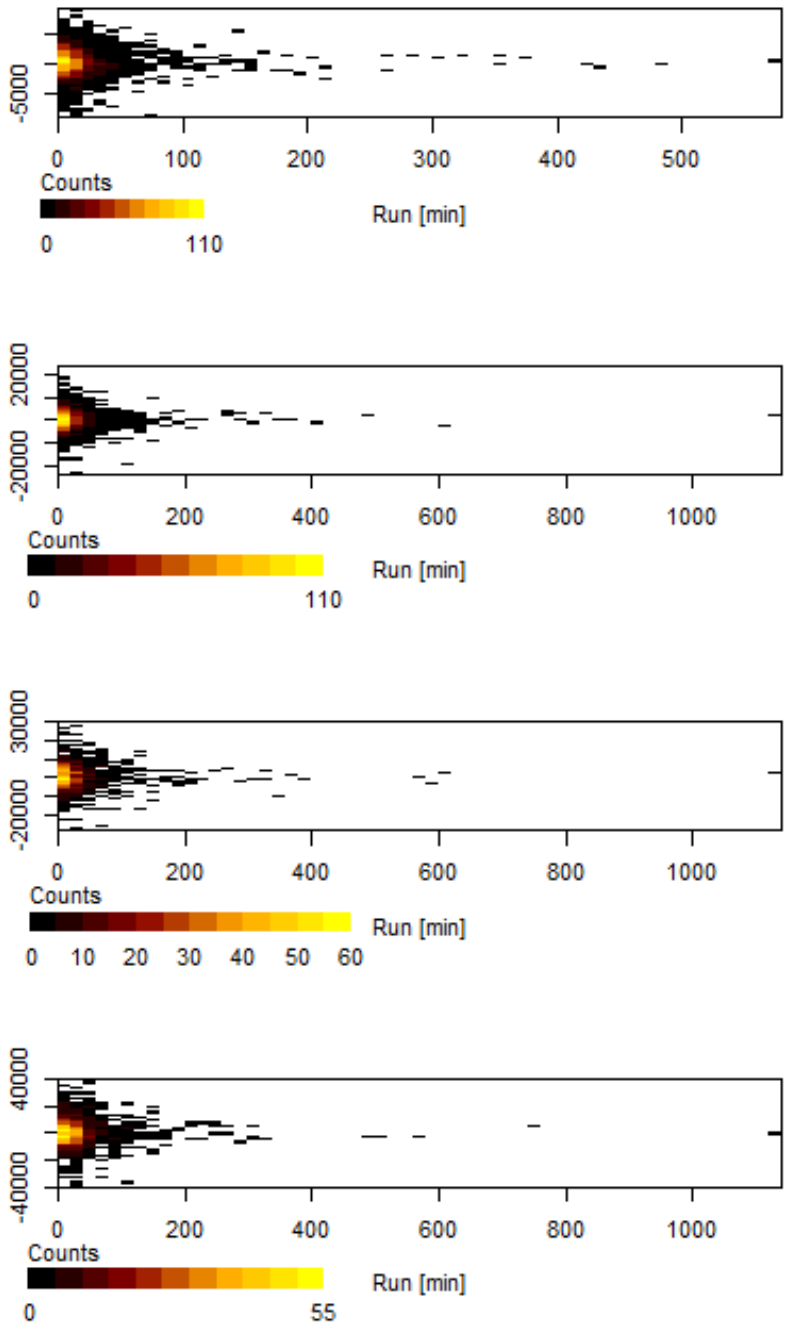

Figure C.8: Bivariate distribution of wind power, ramp rise versus run, as a function of the aggregation level of wind turbines; top subplot is $\epsilon = 25\%$ total wind farm, followed by $\epsilon = 50$, 75, and 100%, respectively. An $\epsilon = 3\%$ of the maximum capacity was used for the month of December.

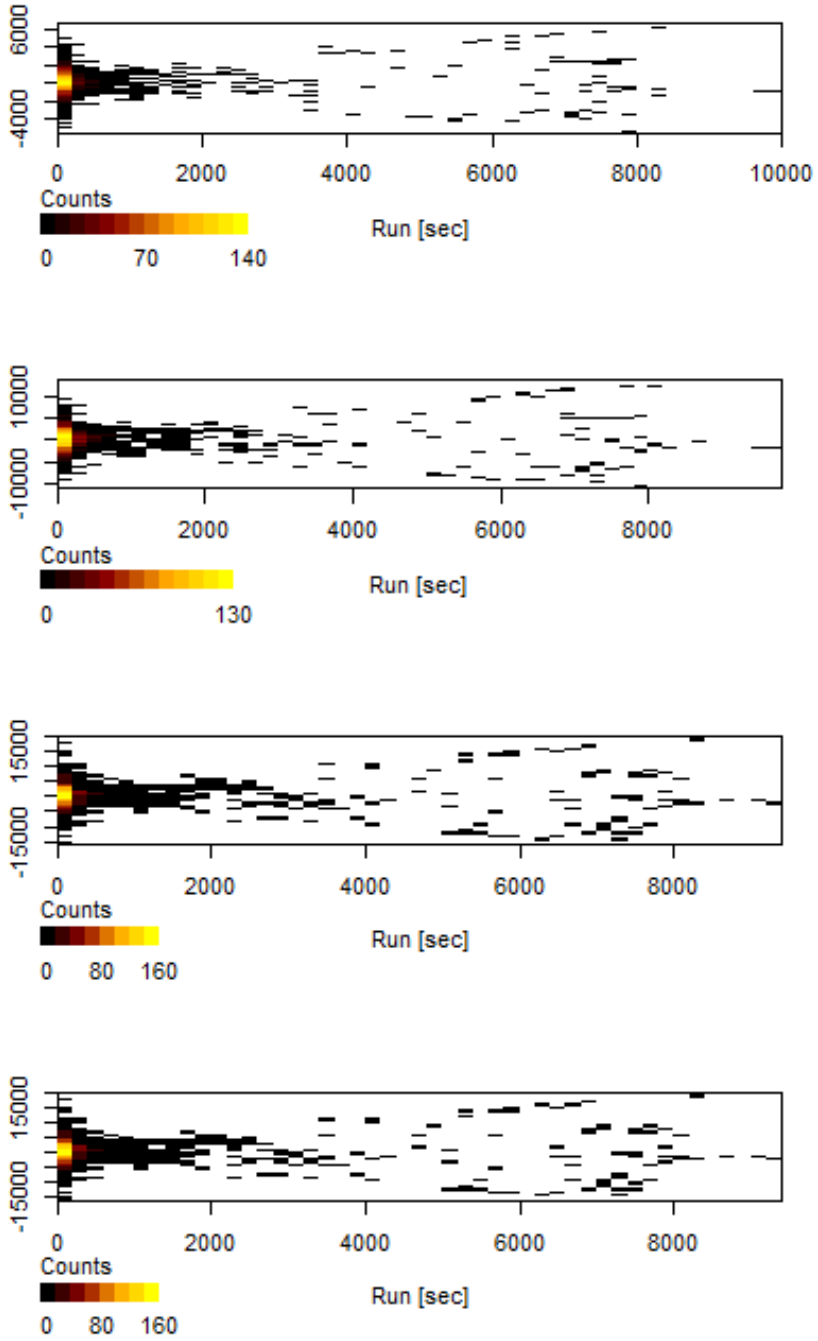

Figure C.9: Bivariate distribution of solar power, ramp rise versus run, as a function of the aggregation level of PV modules; top subplot is $\epsilon = 25\%$ total plant, followed by $\epsilon = 50$, 75, and 100%, respectively. An $\epsilon = 3\%$ of the maximum capacity was used for the month of December

the array, and spatial correlations in power variability diminished with distance. In either case, and as is frequently observed, variability was smoothed with increasing plant (area) size.

It was of interest to determine the extent of smoothing observed in the extracted ramp events. Figures C.8 and C.9 give the bivariate distributions of wind and solar power, respectively, as a function of levels of aggregation in either percentage of wind turbines or PV modules; in both figures, the percentage was, from top-to-bottom subplots, equal to 25, 50, 75, and 100% of the total fleet in the month of December. As shown by the wind power ramp extraction of Figure C.8 (i.e. rise [MW] versus run [min]), there was a slight reduction in the dispersion of ramps as the aggregation level increased. In the solar power ramp extraction of Figure C.9 (rise [kW] versus run [s]), the same trends as the wind example were noted; however, the correlated nature caused by diurnal behavior became more pronounced with increasing levels of aggregation. In addition, the frequency of more immediate (quick) ramps seemed to level off around one-half the total capacity of the PV solar plant.

C.5 Conclusions

The forecasting of solar and wind power ramps is a major area of concern in the field of variable generation forecasting. In this work, the application of a data compression technique to the identification of solar and wind power ramps was shown. Because these ramping events are one of the most pressing concerns of system operators in balancing areas with large penetrations of variable generation, this automated identification process is helpful toward creating algorithms and assessment metrics that can better forecast variable generation ramps and their economic impact. One of the critical issues in wind and solar power forecasting is that the metrics used to assess forecasting techniques are simple statistical measures that do not take into account the factors that are most critical for power system operations. For example, because power systems have means by which they can compensate for small forecast errors, and large forecasting errors are both expensive and can present reliability concerns, it would be better to improve the forecasting for these extreme events, even at the cost of slightly decreased performance during the rest of the times. This is

something that is very difficult to capture with the currently used statistical techniques in which the impact of a large number of small error events can overwhelm the impact of a small number of large error events. Because ramping events comprise a large percentage of these large error events, their automated identification is an important step toward developing metrics that can be used to tune forecasting algorithms to consider their importance. In addition, similar identification techniques could be used actively in system operations. One possible example of how this could be used to improve operations would be an increase in reserves being triggered by the signal when a down ramp in power output had begun. The automated identification would also be useful in assessing probabilistic forecasts. Some system operators currently request that downward ramps in wind power are forecast in a probabilistic manner, in a separate forecast product from the normal forecasts. These forecasts indicate degree-of-belief, giving the likelihood of a down ramp occurring in the specified time frame, and the automated identification techniques advocated here could lead to improvements in assessing system performance.